Applied Ocean Research 145 (2024) 103935

Contents lists available at ScienceDirect

Applied Ocean Research

journal homepage: www.elsevier.com/locate/apor

Gaussian process regression approach for predicting wave attenuation through rigid vegetation

Kristian Ions *, Alma Rahat, Dominic E. Reeve, Harshinie Karunarathna

Swansea University, Swansea, United Kingdom

ARTICLE INFO

Keywords:
Wave attenuation
Coastal vegetation
XBeach
Machine learning

ABSTRACT

Numerical modelling in the coastal environment often requires highly skilled users and can be hindered by high computation costs and time requirements. Machine Learning (ML) techniques have the potential to overcome these limitations and complement existing methods. This is an exploratory investigation utilising a Gaussian Process (GP) data-driven modelling approach that can reproduce, for the given range of conditions in this study, the results of a widely used process-based model, XBeachX, when applied to the challenging problem of wave attenuation through vegetation. This study utilises efficient sampling strategies for data exploration, providing a valuable framework for future studies. The GP model was trained on a synthetic dataset generated using the numerical model XBeachX, which was calibrated using laboratory measurements. Our findings indicate that well-trained ML models can strongly complement traditional modelling approaches, especially in an environment where data sources are increasingly available. We have also explored the underlying interactions of the GP model's input features and their relationship to the model's output through a sensitivity analysis.

1. Introduction

Coastal communities worldwide are becoming increasingly vulnerable to natural disasters, leading to flooding via storm surges and wave overtopping in low-lying areas. These events significantly threaten coastal infrastructure, residents, and local economies. Climate change is predicted to make these events even more frequent and severe in the coming years (IPCC, 2021).

A wide range of solutions is needed to mitigate coastal floods. One promising option is to utilise nature as it is or combine natural solutions with some small-scale management interventions, called Nature-based Solutions (NbS). Research into the benefits of NbS is growing (Pontee et al., 2016; Sutton-Grier et al., 2015). NbS can offer buffer zones to coastal communities during adverse weather events and climate regulation while conserving natural ecosystems, reducing poverty, increasing economic growth, and providing food and livelihoods (Cohen-Shacham et al., 2016). Coastal ecosystems are one type of NbS. Examples include salt marshes in estuaries (McOwen et al., 2017), seagrass beds in sheltered bays (Short et al., 2007), and mangrove forests along coastlines (Giri et al., 2011). These species are crucial for shaping their environment, conserving local ecology, and benefiting local economies (Temmerman et al., 2013; Himes-Cornell et al., 2018). Using coastal vegetation as natural protection has been extensively studied and is widely acknowledged. Multiple field studies have explored wave attenuation through vegetation (Möller et al., 2014; Zhang et al., 2012; Yang et al., 2012; Nardin et al., 2020; Quartel et al., 2007; Jadhav et al., 2013). The consensus on coastal vegetation as a means of coastal defence is that wave energy is dissipated, acting as a natural barrier to coastlines. Coastal vegetation is also an effective relief for tsunamis; Kathiresan and Rajendran (2005) found that during the 2004 Indian Ocean tsunami, mangrove forests acted as a natural barrier, mitigating the impact on the coastline.

In addition to wave attenuation, several field studies have concluded that coastal vegetation reduces the magnitude of storm surges in the surrounding area (Kirwan et al., 2016; Shepard et al., 2011). Numerical modelling studies confirm the damping effect of vegetation on storm surges (Fairchild et al., 2021; van Rooijen et al., 2016; Bennett et al., 2020, 2023). The economic impact of coastal vegetation as a natural buffer zone has been found to reduce flood damage costs by up to 37% across large salt marsh estuaries (Barbier et al., 2011). However, it is worth noting that the observed effects of vegetation on wave attenuation and storm surge reduction can vary widely in the field, with studies reporting attenuation of waves ranging from 10% to 90% and storm surge reduction varying significantly as well (Anderson et al., 2011;

* Corresponding author.
E-mail address: 920039@swansea.ac.uk (K. Ions).

https://doi.org/10.1016/j.apor.2024.103935
Received 13 November 2023; Received in revised form 20 January 2024; Accepted 15 February 2024
Available online 27 February 2024

Möller et al., 2014).

In parallel with filed studies, numerous laboratory investigations (Losada et al., 2016; Hu et al., 2014; Kobayashi et al., 1993; van Veelen et al., 2020; Maza et al., 2015; Ozeren et al., 2014; Koftis et al., 2013) have been carried out to investigate wave attenuation by coastal vegetation. These studies have investigated the complexities of hydrodynamic-vegetation interactions in greater detail, attempting to understand the variability observed in nature, and have broadly reached conclusions similar to those of field-based studies. They have made great insights into the complex governing physics of vegetation-assisted wave attenuation, finding that a wide range of hydrodynamic (e.g. wave height, wave period, water elevation, tidal currents) and plant morphological and mechanical properties (e.g. stem height, stem width, density, rigidity) individually and collectively contribute to wave attenuation.

Numerous theoretical studies focusing on quantifying wave attenuation through vegetation accurately have been reported, each employing a different approach. Initially, vegetation was considered as a viscous layer (Price et al., 1968; Mork, 1996) or as an enhanced bottom drag coefficient (Camfield, 1983). Dalrymple et al. (1984) presents attenuation as a function of incident wave and vegetation parameters. This method has been extensively validated and most commonly used. Dean and Dalrymple (1991) demonstrated the validity of Linear wave theory over an area of simplified rigid cylinders, and the wave energy reduction was related to the work done by the stem drag force. This was further extended to random waves by Mendez and Losada (2004). The dominant issue for these theoretical solutions is the drag coefficient, for which there is no unified methodology of calibration. While several formulae have been successfully applied to estimate the drag coefficient, a unifying model remains absent. This issue is particularly prevalent in numerical models where calibration of the drag coefficient for specific vegetation is only sometimes possible. Although some studies have developed mathematical solutions which eliminate drag coefficient calibration (van Veelen et al., 2021; Maza et al., 2022), they still require further validation against wider conditions before being used widely.

Building on those laboratory studies and theoretical frameworks, numerical models that can capture wave attenuation on vegetation have been developed, including Bousinessq-type (Augustin et al., 2009), Reynold-averaging Navier Stokes (RANS) equations (Li and Yan,2007; Maza et al., 2013) and shallow water equations (Wu and Marsooli, 2012). The numerical schemes for rigid vegetation have been incorporated into existing wave models, such as SWAN (Suzuki et al., 2012), SWASH model (Suzuki et al., 2019), XBeach-Veg model (van Rooijen et al., 2016), CSHORE (Zhu et al., 2018) and WWIII (Abdolali et al., 2020).

The development of numerical models has aided the simulations of vegetation-assisted wave attenuation in coastlines, thus helping engineers, scientists, and policymakers make informed decisions on nature-inspired coastal defence solutions. Additionally, they can generate vast synthetic datasets under varying environmental conditions which can be used to understand wave attenuation problems under a wide range of environmental conditions. However, numerical models can be computationally expensive and time-consuming. They require expertise to set up models for the site concerned and ensure the model is accurately calibrated and adequately validated. This trade-off between model accuracy and simulation duration and costs can limit the utility of such models.

Machine learning (ML) has emerged as a promising alternative or supplementary approach to numerical modelling and has proven the potential to address certain limitations of numerical modelling (Panchigar et al., 2022). The rapidly expanding field of ML and ever-increasing data availability present new opportunities to develop improved predictions. This has led to increased interest in applying ML techniques in the context of coastal systems (Goldstein et al., 2019). ML algorithms adapt to data during the training process. This can be done without requiring expert knowledge. Insights and predictions that were not previously obvious can be extracted.

ML techniques are believed to be poised to revolutionise the field of engineering by using data-driven approaches to complex model systems (Molnar, 2020). ML has emerged as a powerful tool for analysing complex and nonlinear relationships in data. This is achieved by using specific algorithms that can be learned from new data, which allows the analysis of complex nonlinear relationships (Salehi and Burgueño, 2018). The two primary types of ML problems are unsupervised and supervised Learning. In unsupervised learning, models are trained to predict results by identifying patterns in data using methods such as clustering and density estimation. In contrast, in supervised learning, trained models approximate a function between variables and output and apply this function to unseen data to make predictions for continuous output (in regression) or discrete class labels (in classification). Examples of supervised Learning methods include Naïve Bayes, Support vector machines, Random Forest, Decision Trees, Linear Regression, Logistic Regression, Neural Networks, and Gaussian Processes Regression. In addition, reinforcement learning is a distinct ML paradigm from these two, where the goal is to interactively learn effective policies for controlling an environment (Sutton and Barto, 2018).

ML has already been successfully applied in coastal modelling and engineering. For instance, supervised ML methods have been applied to ocean wave modelling by James et al. (2018) and Minuzzi and Farina (2023), who trained their respective ML models on datasets generated using physics-based wave models. Both studies concluded that their predictive performances were comparable with the physics-based wave models. Gracia et al. (2021) and Wang et al. (2021) used a dataset of wave height measurements to train their ML models for nearshore wave predictions. Their results showed that the ML models outperformed traditional methods of wave height estimation, indicating that ML techniques can help improve the accuracy of wave height predictions. Furthermore, large datasets were used to forecast wave overtopping more accurately than existing empirical formulae, using ML techniques (den Bieman et al., 2021; Hosseinzadeh et al., 2021).

In wave and flow attenuation, Kim et al. (2022) demonstrated two applications of ML. Firstly, they predicted wave attenuation over an artificial reef using an artificial neural network (ANN) model trained on hydraulic experimental data, which yielded high accuracy. Secondly, they showed the power of ML models to perform sensitivity analysis on complex systems. Wang et al. (2023) also demonstrated the ability of ANNs to predict the drag coefficient for rigid vegetation. Wang et al. (2021) successfully implemented a genetic programming routine to derive a new predictor for the drag coefficient of flexible vegetation. Tinoco et al. (2015) used a genetic programming routine to derive a physically sound equation relating flow and vegetation characteristics to depth-averaged velocity over submerged rigid cylinders. However, the study highlighted the importance of experts in selecting the best numerical and physical solutions. Conversely, Maji et al. (2022) showed that the system could automatically learn without explicitly being programmed by using the polynomial regression ML method. This reveals a disadvantage of ML, with each ML application requiring individual assessment, and the most suitable approach must be selected based on the training data type and the target data. Several other studies have provided in-depth reviews of the application of ML techniques in coastal processes and modelling, highlighting the accuracy and performance advantages of ML over traditional empirical approaches (Chau, 2006; Hsieh, 2009; Valentine and Kalnins, 2016; Dwarakish and Nithyapriya, 2016; Goldstein et al., 2019; Beuzen et al., 2019). In most cases, the ML performance exceeded the traditional empirical approach or achieved comparable predictions with numerical models.

A multitude of successfully implemented ML applications have been discussed above. The successful application of ML in wave height prediction, including wave modelling and flow attenuation, highlights the potential for machine learning techniques to improve accuracy and efficiency in various coastal engineering applications. However, it is essential to consider the specific needs of each application and select the

most appropriate ML approach accordingly. As previously discussed, ML often requires large datasets, which can be a disadvantage, due to limited publicly available data. Data sparsity can lead to underperforming models, larger model uncertainty, model bias or render ML unusable, especially when there is little control over what data is being collected. In this study to bypass this issue, we employ a process-based model XBeachX, through which a synthetic dataset of any size can be created to train the model, albeit with the associated computational expense. Additionally, the outputs from ML must be examined by a professional to ensure the predictions align with the physical requirements of the studied system.

This paper takes the first steps of applying an ML approach to model wave attenuation through rigid, submerged and emergent coastal vegetation using Gaussian Processes (GP). We also aim to investigate the relationships between the underlying input parameters and the target variable. Despite the increasing popularity of ML methods in coastal science and engineering disciplines, GPs are not yet extensively explored to model coastal processes. They offer several advantages over other ML techniques, such as providing probabilistic predictions, and a principle framework for incorporating prior knowledge. Furthermore, GPs can deal with nonlinear relationships between variables and missing or noisy data. In addition, coastal engineers and practitioners often require a measure of uncertainty for given predictions for them to make informed decisions on different coastal management choices and assess risks, which is a key feature of GP modelling. We believe that our approach will provide valuable insights into the problem of wave attenuation in coastal vegetation and demonstrate the efficacy of GPs as a powerful tool for modelling complex datasets.

This study will also contribute to the growing literature on ML applications to coastal modelling and engineering by introducing a flexible, data-agnostic framework based on the GP approach. In the absence of an extensive dataset, we will train the GP model on a synthetic numerical simulation dataset generated using the XBeachX coastal model. XBeachX was calibrated and validated using a set of experimental data. The use of XBeachX, allows the creation of a reliable, physically accurate, synthetic dataset from which we can test our hypothesis – whether a GP approach can emulate a process-based model in predicting wave attenuation through vegetation. The authors want to emphasise here XBeachX was initially calibrated using a limited set of hydrodynamic conditions and specific plant characteristics, making this model unsuitable at this time for use outside these conditions The GP model predicts wave attenuation over rigid, emergent, and submerged vegetation while accounting for uncertainty, which is often unaccounted for in previous studies which have incorporated various ML tools. In our approach, we will address four issues, which have not been dealt with before:

(i) A suitably trained ML model would provide a rapid assessment of wave attenuation in a matter of seconds as opposed to numerical models which calculate over minutes to hours. This would reduce computational time, whilst also providing a simple-to-use method, compared to the high skill approach of numerical modelling.
(ii) Providing uncertainty for point predictions is vital for design consideration, which is made possible through the probabilistic approach of GP regression.
(iii) Providing a methodological framework that can be applied to a broader range of coastal engineering uses.
(iv) Exploring the input parameter space using existing sensitivity analysis methods to understand the correlations of the GP and determine if those methods can provide an insightful analysis to the key drivers of vegetation-assisted wave attenuation relating to theoretical understanding.

This paper is structured as follows: Section 2 provides the theoretical background on wave attenuation over rigid vegetation and background on the GP method. Section 3 describes the methodology, including a brief overview of the numerical model XBeachX, its calibration and validation, and the experimental set-up. Section 4 explains the development of the GP model. Section 5 presents and discusses the results of our GP modelling. Finally, the paper ends with conclusions presented in Section 6.

2. Theoretical background

2.1. Wave attenuation through rigid vegetation

In general terms, assuming plant geometry to be a rigid, vertical cylinder, wave energy is dissipated over a length of vegetation patch is controlled by the conservation of wave power (Dalrymple et al., 1984). For a flat bed and monochromatic wave approach,

$$\boldsymbol{C}_g = \frac{\partial \boldsymbol{E}}{\partial \boldsymbol{x}} = -\boldsymbol{n}_v \in_v \tag{1}$$

where $E = \frac{1}{8}\rho g H^2$ is energy content of a monochromatic wave, $c_g = \frac{\omega}{2k}\left(1 + \frac{2kh}{\sinh(2kh)}\right)$ is wave group velocity where $\omega = 2\pi/T$ is wave angular frequency, $k = 2\pi/\mathrm{L}$ is wave number, h is the total water depth and n_v is the number of stems/m^2. The energy dissipated per stem is given as

$$\in_v (\boldsymbol{x}) = \int_{s=0}^{h_v} \overline{\boldsymbol{F_W}\, \boldsymbol{u}} \boldsymbol{ds} \tag{2}$$

where $\mathrm{F_W}$ is the total wave force in the x-direction (Morison et al., 1950), u is the local flow velocity, and the overbar represents the wave's phase averaging and h_v is the stem height.

$$F_w = \frac{1}{2}\rho C_D b_v n_v |u| u \tag{3}$$

in which C_D is the drag coefficient, b_v is the stem diameter and ρ is density of water. Dalrymple et al. (1984) showed that by assuming linear wave theory to be valid, then waves decayed reciprocally as they propagated through vegetation, which can be expressed by

$$H = \frac{H_0}{1 + \beta x} \tag{4}$$

where H_0 is the approaching wave height at the upstream edge of the vegetation patch and β is the wave damping coefficient, with x representing cross-shore distance. β is expressed by

$$\beta = \frac{4}{9\pi} C_D b_v n_v H_0 k \frac{\sinh^3 k h_v + 3\sinh k h_v}{(\sinh 2kh + 2kh)\sinh kh} \tag{5}$$

where $a = h_v/h$ is the submergence ratio.

2.2. Gaussian processes regression as a machine learning technique

A GP is a non-parametric, Bayesian regression approach, that can infer a probability distribution over all possible functions that may be able to model the observed data. Its ability to provide predictions and the uncertainty of the predictor via the posterior probability density are crucial factors that make GP models a preferred choice over other machine learning approaches. Furthermore, GP models typically require less data than other approaches, making them useful for small sample sizes.

The method works by first assuming a GP prior, usually with zero mean and unit variance. This means before any data has been observed, the prediction for any independent variable vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ where x_i is the ith component would be a Gaussian distribution over the function output, i.e. $f(x) \sim \mathcal{N}(0, 1)$. Another prior in-

formation that is considered is the Kernel or Covariance function $\kappa(x', x'', \boldsymbol{\theta})$ that models how much the output function responses vary between any two input vectors x' and x'' controlled by its hyperparameters θ. There are many choices for the Kernel function that could be used, for example, radial basis function (RBF), rational quadratic (RQ), Matern, etc., and it is also possible to combine kernels together (Duvenaud, 2014).

Once we observe the outputs of m input vectors, we construct a dataset with a matrix $X = (x_i^j)_{i=1,j=1}^{i=n,j=m}$ and the associated output vector $f = (f^j)_{j=1}^{j=m^\top}$. Next, using this data we train the GP, which constitutes locating the covariance function hyperparameters θ^* that maximises the marginal likelihood of the data (Williams and Rasmussen, 2006):

$$\log p(X,f|\theta) = -\frac{1}{2}\log|K| - \frac{1}{2}f^\top K^{-1}f - \frac{N}{2}log(2\pi) \tag{6}$$

where K is the covariance matrix with each element $K_{c,d} = \kappa(x^c, x^d)$ such that $x^c, x^d \in X$.

Now, the posterior distribution over the function space for an arbitrary x is entirely defined through a mean function $\mu(x \mid \theta^*)$ and a variance function $\sigma^2(x \mid \theta^*)$, i.e. $p(f|x, \theta^*) \sim \mathcal{N}(\mu(x|\theta^*), \sigma^2(x|\theta^*))$; for notational simplicity we exclude θ^* henceforward. These functions are defined as:

$$\mu(x) = f^\top K^{-1}\mathrm{k} \tag{7}$$

$$\sigma^2(x) = \kappa(x,x) - \mathrm{k}^\top K^{-1}\mathrm{k} \tag{8}$$

where, $k = (k_1, \ldots, k_m)^\top$ with $k_l = \kappa(x, x^l)$ is the covariance between x and lth input vector $x^l \in X$.

A simple 1D case is demonstrated in Figs. 1 and 2 to understand the GP better. For instance, the target function can be defined as a sine function. The data is generated and pre-processed. In this example, a Latin-hyper-cube sample (LHC) generates the query data, and the training data consists of four points from [0, 1]. LHC sampling is a statistical technique that ensures representative sampling across the input parameter space by dividing the variable range into equal intervals and selecting one value from each interval, developed by McKay et al. (1979). The test dataset is comprised of 100 points from [0 1], and the RBF kernel is selected. The GP model is subsequently trained on the training dataset, and during this training, the model optimises the hyperparameters over a user-defined number of iterations. In this example, the number of iterations is 100. The trained model is then evaluated against the test data over the sin function. Fig. 2 demonstrates that there is some mean prediction and associated uncertainty for each sampled test point. The model produced zero error at locations where training data was available encapsulating the assumption that there is no measurement error; the model had a large error for locations far away from trained data. The reader is referred to Williams and Rasmussen (2006) for a more detailed and comprehensive explanation.

2.3. Efficient sampling strategies for data exploration

In this study, we explored diverse sampling or data collection strategies while training the GP model. In particular, we investigate the following methods.

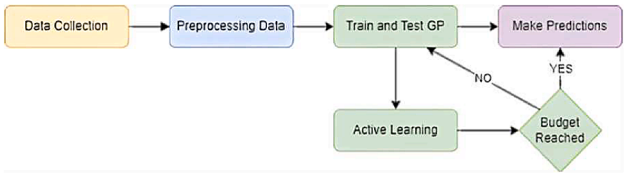

Fig. 1. Simple flowchart for a GP with optional Active Learning (AL inclusion). The budget is user-defined.

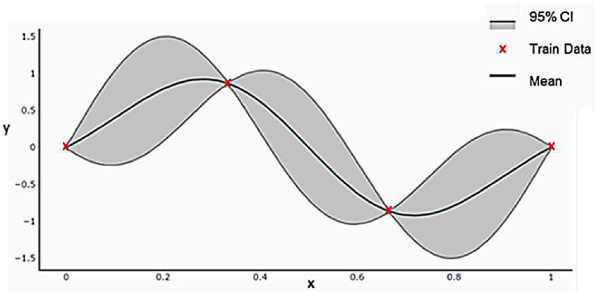

Fig. 2. Example case of GP on a sine function using RBF kernel.

- Random sampling, the simplest method, involves randomly selecting data points.
- Latin-hypercube sampling (LHC) is a statistical approach to selecting data to improve sparse coverage of input space.
- Active Learning (AL) is a sequential and iterative approach that actively selects the most informative data points, attempting to reduce model uncertainty and enhance performance at every iteration.

Fundamentally, the greater the number of data points, with the appropriate coverage of the input space, the smaller the uncertainty of the fitted function. Fig. 2 above indicates where data is sparse, larger uncertainties are estimated. In the case of LHC and Random sampling, increasing the number of samples would result in an expected reduction in uncertainty. For AL, new data points are selected at the regions of maximum uncertainty, and therefore, the GP predictions are likely to reduce predictive uncertainty whilst using fewer data points.

AL is the branch of ML which can address the issue of optimum data acquisition whilst limiting the need for user input. Previously, we demonstrated that a GP can be fitted for a data set of N number input features (N). However, there are regions of high uncertainty in Fig. 2. An AL algorithm selects a new point at the maximum uncertainty, and evaluate the output. This is then followed by a retraining of the GP model where the training dataset is augmented with the newly evaluated point. This approach aims to make informed decisions and promote data-efficient learning in a manner that confers better performance when compared to random sampling (Kingma and Ba, 2014). This is particularly useful where time or data accessibility is an issue – such as computationally expensive numerical models or limited laboratory time. It should be noted that locating the maximum uncertainty solution for AL is an optimisation problem itself with an optimal solution: $x^* = \underset{x \in \mathscr{X}}{\arg\max}\, \sigma^2(x)$, where $\mathscr{X}$ is the space of all possible input vectors.

The Covariance Matrix Adaptation Evolution Strategy, CMA-ES, was implemented in this study to identify the regions of maximum uncertainty. The CMA-ES is a second-order optimisation approach. It is highly suited for local and global optimisation problems and has been shown to have superior performance compared to other popular choices of algorithms (Hansen and Ostermeier, 2001; Hansen and Kern, 2004; Hansen, 2009).

An example of AL, using the CMA-ES algorithm, is demonstrated briefly below in Fig. 3. Taking the previous example in Fig. 2 for a sin function trained on an RBF kernel and running three iterations of AL, the overall predictive uncertainty is reduced, where it was previously largest. Thus, the GP model's performance is quickly improved. The AL algorithm is summarised in Fig 7.

The three sampling methods are used to develop the GP model and are compared in Section 4.

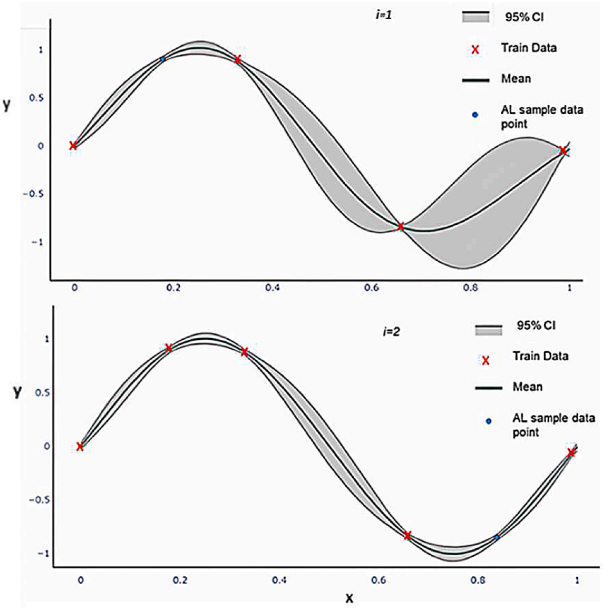

Fig. 3. An example illustration of AL. Top figure shows the fitted GP from Fig. 2, with an additional sample point (blue cross) from AL ($i = 1$). The bottom figures shows the same again with a new AL sample point ($i = 2$).

2.4. Sensitivity analysis of ML model to explore the input parameter space

Primarily, ML models are correlation models, and care must be taken to avoid casual relations being inferred from datasets (Arjovsky et al., 2019). They require oversight and guidance to avoid spurious correlations (Vigen, 2015). Whilst ML models can produce highly accurate predictions, the process is often considered a 'black box', with the hidden connection between target data and input parameters. Transparency and interpretability of ML models are vital. In many science and engineering disciplines understanding how individual components contribute to target outputs aids in a better understanding of the studied processes. Additionally, Molnar (2020)speculated that transparency will boost the adoption of such techniques.

Whilst the application of ML continues to expand, there must be an emphasis on the interpretability of the ML models (Doshi-Velez and Kim 2017). Understanding what is happening behind the black box provides insight into the modelled processes. This can be beneficial in improving model performance and explaining outcomes, and it also reinforces trust in the ML methods (Molnar, 2020; Ribeiro et al., 2016).

Here, we perform a sensitivity analysis on the GP model input parameter space. Several options are available; however, this study only explores the SHapley Additive exPlanations (SHAP) package (Lundberg and Lee, 2017). The package implements the theory first developed by Shapley (1953) relating to game theory. Originally a method used for assigning payouts to players depending on their contribution, the underlying principle has been adapted to ML by Lundberg and Lee (2017). The Shapley value can be calculated by considering all possible combinations of input variables and measuring the value each input variable contributes to the target output variable.

The Shapley value is defined as a value function (V) for the subset of input variables, S. The contribution of the input variable is weighted and then summed over all other input variable combinations (Molnar, 2020). The Shapley value is expressed mathematically as

$$\phi_j(S) = \sum_{S\subseteq\{1,\ldots,p\}/\{j\}} \frac{|S|!(p-|S|-1)!}{p!}\left(V(S\cup\{j\}) - V(S)\right) \tag{9}$$

where x is the input parameters to be explained and p is the number of features. $V(S)_x$ is the prediction of the parameter values in S. These are marginalised over all other parameter values not in S.

$$V_x(S) = \int \widehat{f}(x_1,\ldots,x_p)d\mathrm{P}_{x\notin S} - E_X(\widehat{f}(X)) \tag{10}$$

The SHAP package and other interoperability algorithms can be utilised to study global and local interoperability. Global interoperability benefits engineering by allowing a comparative analysis of input parameter variables across different study methods, i.e., experimental, numerical and ML. The importance of an input variable on the target variable can be deduced, which can then be combined with theoretical knowledge. Once we understand how the underlying input variables interact, this can lend confidence to ML outputs, especially when they co-align with existing knowledge. Local interoperability can offer cases-specific insights, which can be beneficial when more thorough analysis is required.

3. Generation of training and testing data

3.1. Numerical model

ML techniques require sufficient data to derive meaningful predictions of the target process with an acceptable level of uncertainty. In the absence of a suitable field or experimental dataset of wave attenuation on coastal vegetation, we used the numerical model XBeach-X, which is a widely used open-source coastal hydro-morphodynamic model. XBeach-X is an amalgamation of XBeach (Roelvink et al., 2009), which was developed initially to simulate dune erosion of sandy beaches due to hurricanes in which waves were resolved at wave group scale and XBeach-G developed for gravel beach applications (McCall et al., 2014, 2015). XBeachX was also updated with vegetation dissipation effects by van Rooijen et al., 2016. Three wave propagation modes are available in XBeach, namely stationary, surfbeat and non-hydrostatic. This study uses XBeachX Stationary mode wave modelling. This best simulates sinusoidal, non-varying wave heights of regular and irregular oscillatory waves, similar to the waves produced under laboratory conditions (Section 3.1). Stationary mode is suitable for shorter wave motions, neglecting infragravity waves, and models wave motions using the HIWSA equations (Holthuijsen et al., 1989). The model resolves wave propagation based on the short-wave action balance equation:

$$\frac{\partial A}{\partial t} + \frac{\partial c_x A}{\partial x} + \frac{\partial c_y A}{\partial y} + \frac{\partial c_\theta A}{\partial \theta} = -\frac{D_w + D_f + D_v}{\sigma} \tag{11}$$

where D_w wave-breaking dissipation, D_f is bottom friction dissipation and D_v is dissipation due to vegetation. c_x, c_y and c_θ are wave action propagation speeds. Wave action A can be calculated as

$$A(x,y,t,\theta) = \frac{S_w(x,y,t,\theta)}{\sigma(x,y,t)} \tag{12}$$

where θ represents the angle of incidence for the x-axis, S_w represents the wave energy density in each directional bin and σ the intrinsic wave frequency.

In stationary mode D_w is calculated using the equation proposed by Baldock et al. (1998) for wave dissipation.

D_f is calculated as

$$D_f = \frac{2}{3\pi}\rho f_w\left(\frac{\pi H_{rms}}{T_{m01}\sinh kh}\right)^3 \tag{13}$$

where H_{rms} is root-mean-squared wave height, T_{m01} is mean spectral wave period, k is wave number, h is water depth, ρ is fluid density and f_w

is short-wave friction coefficient which is a user calibrated parameter.

The vegetation dissipation is based on the modelling approach of Mendez and Losada (2004), which was adapted in XBeach by van Rooijen et al., 2016. The model allows the implementation of multi-layered vegetation represented through several parameters. The dissipation is represented by

$$D_v = A_v \frac{\rho C_D b_v N_v}{2\sqrt{\pi}} \left(\frac{kg}{2\sigma}\right)^3 H_{rms}^3, \tag{14}$$

$$A_v = \frac{\left(\sinh^3 k\alpha h - \sinh^3 k\alpha h\right) + 3\left(\sinh k\alpha h - \sinh k\alpha h\right)}{3k\cosh^3 kh}$$

where C_D is a (bulk) drag coefficient, b_v is the vegetation stem diameter, N_v is the vegetation density per m^2, and α is the relative vegetation height (h_v/h). Vegetation is considered as rigid cylindrical stems.

Only intermediate-depth conditions were considered for the scope of this study. As a result, wave breaking did not occur, and the roller energy balance and subsequent shallow water equations are not applicable. The study solely focused on wave dynamics through vegetation therefore, the sediment transport and morphology capabilities of XBeachX were turned off for all simulations. The drag coefficient is user-calibrated and not updated within the model, which can lead to modelling inaccuracies. Wave non-linearity is not accounted for in the HIWSA equations either.

XBeachX implements vegetation using a binary grid system imposed upon the numerical grid. The length of the vegetation, L_v, N_v, and the height of the vegetation, h_v as well as C_D are user-defined.

In this study, we set up a numerical wave flume using XBeachX, which is similar to the laboratory wave flume of Swansea University, UK. The rationale for performing laboratory-scale numerical experiments is that the model can be calibrated and validated using the experimental data generated in the laboratory wave flume. We can also generate controlled numerical experiments of vegetation-assisted wave attenuation under a wide range of input conditions.

3.2. XBeachX calibration and validation

XBeachX model was calibrated and validated using experimental data collected at Swansea University Coastal Engineering Laboratory wave flume, which is 30.7 m long, 0.8 m wide and 1.2 m deep. The experimental set up is shown in Fig. 4. The experiments measured wave attenuation on a vegetation patch under a range of hydrodynamic conditions using a series of wave gauges. The experimental results pertaining to rigid vegetation, which has similar flexural rigidity to common saltmarsh species (van Veelen et al., 2020), is used in this study.

The water levels were selected to represent emergent and submerged vegetation canopies. The water level of the flume was kept constant for the duration of each experimental run. The wave heights were recorded by three wave gauges (WG); WG1 was located 1.05 m upstream from the vegetation patch and was used as an estimate of pre-vegetation wave height, H_{start}; WG2 was located in the centre of the vegetation patch. Lastly, WG3 was located 0.1 m downstream of the vegetation patch, which was used as post-vegetation wave height, H_{post}. The input wave height H_0, was taken as the wave height generated by the paddle. The waves considered for analysis were selected when the incoming waves reached 95% of the significant wave height (H_s). Whilst the wave damper dissipated most of the waves, a small amount (< 10%) of reflection remained. Therefore, waves measured after the reflected wave reached WG3 was disregarded to avoid reflection contamination. An extensive account of the experimental set up and the measurement programme can be found in van Veelen et al. (2020).

Scaling issues can be neglected as van Veelen et al. (2020) ensured wave-vegetation interactions were consistent with nature by following previous methods for scaling of vegetation and wave conditions (Ghisalberti and Nepf, 2002; Luhar and Nepf, 2016; Luhar et al., 2017; Zhang and Nepf, 2021). The following non-dimensional parameter was identified to ensure the model conditions were representative of real-world conditions; kh (ratio of wavelength to water depth); L_b/h (ratio of blade length to water depth); H/h (wave height to water depth);Froudes number, $Fr = u/\sqrt{gh}$; the vegetation Reynolds number, $Re = uB_w/\nu$, and Keulegan-Carpenter, $KC = uT/B_w$, which is the ratio of wave excursion and stem diameter, which is a predictor for drag coefficients of cylinders (Keulegan and Carpenter, 1956). All cases had a vegetation stem height (h_v) of 0.3 m; stem density (N_v) of 1111 stems/m^2 and blade width (b_v) 0.005 m. In total 23 cases were run for rigid mimics. These cases consisted of; H_s [0.08 m – 0.2 m]; wave period, ET, [1.4 s – 2.0 s] and h [0.3 – 0.6] and can be found in Table 1.

A 1D horizontal numerical wave flume, which replicates the experimental set-up of van Veelen et al. (2020) was created in XBeach. The numerical flume was 30.7 m long and had an equidistance grid with dx = 0.3 m grid spacing.

A generating-absorbing boundary was defined at the paddle and end of the numerical flume. The left and right boundary conditions implement zero velocity at the lateral boundary. The model had an initial spin-up time of 300 s.

The experimental cases in Table 1 are divided as XBeachX model calibration cases, which were selected as the experimental cases representative of the most and the least wave attenuation, i.e. - cases 1 and 22. Five cases (cases 2,8,12,15,19) representing a wide range of environmental conditions were randomly selected for model validation. Lastly, the remaining 15 cases were used for GP experimental validation dataset.

Initially, the calibration cases were run with default XBeachX settings. The hydrodynamics of the XBeachX model were calibrated by

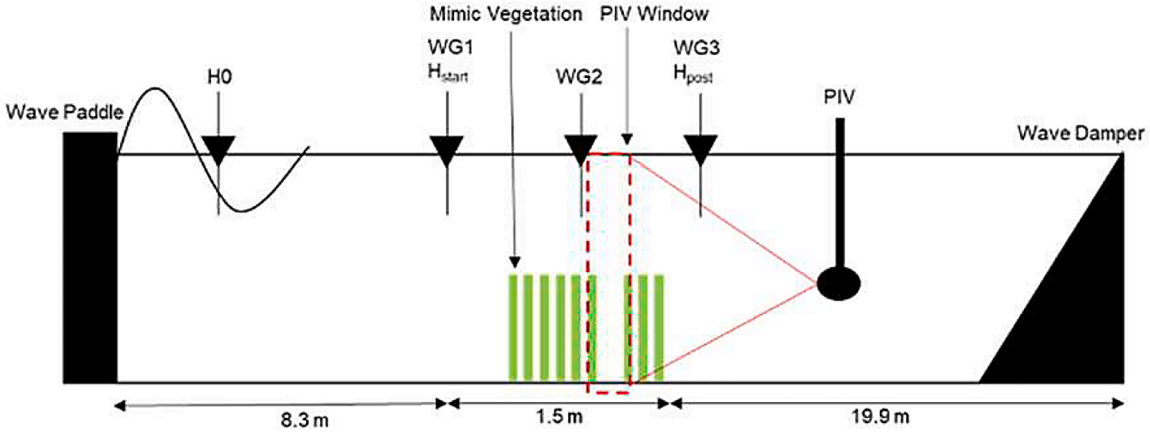

Fig. 4. Schematic of the experimental set-up of van Veelen et al. (2020).

Table 1
Hydrodynamic and vegetation parameters used in physical experiments used for XBeachX calibration and validation.

Case	Use for	hv [m]	Nv [stems/m2]	Lv [m]	h [m]	H_0 [m]	T [s]	H_{lab}/H_0	H_{xb}/H_0
1	Calibration	0.3	1111	1.5	0.6	0.15	1.4	0.92	0.92
2	Validation	0.3	1111	1.5	0.6	0.15	1.6	0.91	0.90
3	GP Test	0.3	1111	1.5	0.6	0.15	1.8	0.90	0.88
4	GP Test	0.3	1111	1.5	0.6	0.15	2.0	0.87	0.88
5	GP Test	0.3	1111	1.5	0.6	0.10	1.8	0.91	0.92
6	GP Test	0.3	1111	1.5	0.6	0.20	1.8	0.86	0.85
7	GP Test	0.3	1111	1.5	0.5	0.15	1.4	0.84	0.86
8	Validation	0.3	1111	1.5	0.5	0.15	1.6	0.83	0.84
9	GP Test	0.3	1111	1.5	0.5	0.15	1.8	0.81	0.83
10	GP Test	0.3	1111	1.5	0.5	0.15	2.0	0.82	0.82
11	GP Test	0.3	1111	1.5	0.5	0.10	1.8	0.85	0.88
12	Validation	0.3	1111	1.5	0.5	0.20	1.8	0.76	0.79
13	GP Test	0.3	1111	1.5	0.4	0.15	1.4	0.79	0.76
14	GP Test	0.3	1111	1.5	0.4	0.15	1.6	0.79	0.75
15	Validation	0.3	1111	1.5	0.4	0.15	1.8	0.75	0.74
16	GP Test	0.3	1111	1.5	0.4	0.15	2.0	0.67	0.73
17	GP Test	0.3	1111	1.5	0.4	0.10	1.8	0.79	0.81
18	GP Test	0.3	1111	1.5	0.3	0.10	1.4	0.75	0.68
19	Validation	0.3	1111	1.5	0.3	0.10	1.6	0.72	0.68
20	GP Test	0.3	1111	1.5	0.3	0.10	1.8	0.74	0.68
21	GP Test	0.3	1111	1.5	0.3	0.10	2.0	0.61	0.67
22	GP Test	0.3	1111	1.5	0.3	0.08	1.8	0.77	0.72
23	Calibration	0.3	1111	1.5	0.3	0.12	1.8	0.69	0.64

comparing the output from WG1 ($H_{post,lab}$) and the same location in the numerical domain ($H_{post,XB}$). The parameters were then optimised through iteration until $H_{post,XB}$ was within 10% of $H_{post,lab}$. As wave breaking is not included, the wave action balance (equation 12) has only dissipation mechanisms, dissipation due to bottom friction and vegetation.

Firstly, the calibration parameter involved in wave dissipation from bottom friction, *fw,* was set at $f_w = 0.075$. Secondly, dissipation due to vegetation was calibrated via the drag coefficient, C_D. Numerical coastal models, such as XBeach, require the calibration of C_D. Currently, there is no effective method for estimating C_D for a range of different hydrodynamic and vegetation conditions. Since the primary objective of this study is to explore the performance of the GP method for predicting wave attenuation through coastal vegetation, we selected a constant value of C_D. Through calibration, C_D=1.25 was selected to best represent the range of conditions used in this study.

The final calibrated model was then used to simulate the five validation cases. These validation results are shown in Fig. 5. The coefficient of determination, R^2, was calculated as 0.85 and the root mean squared error, RMSE, as 0.0027 m, which indicates the fact the model can successfully simulate wave attenuation through vegetation within the limits of the parameter range considered for calibration and validation.

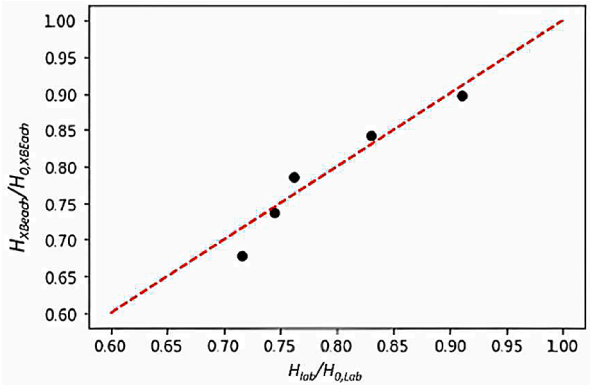

Fig. 5. Actual (experimental wave heights, $H_{lab}/H_{0,lab}$) vs predicted plot (XBeachX predicted wave heights, $H_{XBeachX}/H_{0,XBeachX}$), displaying the validation cases.

3.3. Generation of the synthetic dataset using XBeachX

The calibrated and validated XBeachX wave model was used to generate training testing and validation data for the GP model of vegetation-assisted wave attenuation. The range of input conditions were consistent with the hydrodynamic conditions used in calibration and validation.

i) Training Dataset: The three sampling strategies described in Section 2.3 were used to generate the dataset on which the GP model is trained. Each method generated ten different batches of data to train the model. For each batch of data, a total of 127 samples were generated. The number of input variables was n. In the case of AL the training dataset had an initial sample size (Ni) was $n + 1$ i.e., 7. This was then increased incrementally until the budget (B) was exhausted, where $B = 20$*n, i.e., 120. Therefore, total sample size is Ni+*N*.
ii) Testing Data: Five batches of 100 samples were generated for each batch of testing data. LHC sampling was used to generate the input parameters for the 100 cases.
iii) Validation Dataset: This consisted of two different datasets. Firstly, the remaining experimental data not used in calibrating and validating the XBeachX model, seen in Table 1, was used to validate the final trained GP model. Secondly, a randomly generated XBeach dataset to explore the spatial variability in the prediction from GP, given in Table 3.

Two axioms are considered when developing the GP model:

1) The choice of input features must be sufficient to capture the process accurately.
2) Whether the GP model predictions align with existing knowledge, i.e., $H_{post} > H_{start}$ would not be acceptable since vegetation has been extensively shown to decrease wave height magnitude.

The input parameter space should consist of relevant parameters that characterise the hydrodynamics and vegetation properties to accurately explain the observed trends in wave attenuation. This study's parameter space consisted of essential hydrodynamic and vegetation characteristic parameters. These were identified through literature (Nepf, 2012;

Anderson et al., 2011) and using Eq. (14). Specifically, we identified the primary vegetation parameters as stem diameter, b_v, stem density, N_v, stem height, h_v, and length of vegetation, L_v. Additionally, we considered the key hydrodynamic parameters as wave period, T;wave height at the start of the vegetation patch H_{start} was selected, over H_0, to eliminate the influence of bottom friction dissipation in the predictions; and water depth h. Submerged and emergent conditions are accounted for. The GP model remains simple by keeping the number of input parameters at a minimum. The input parameter space is six-dimensional ($n = 6$) consisting of h_v *[0.*3 m – *0,*6 m*]*, N_v *[333 – 1111]*, L_v *[0.5m-1.*5 m*]*, h *[0.*3 m – *0.*6 m*]*, H_{start} *[0.*08 m - *0.*2 m*]*, and T*[*1 *s* - 2 *s]*. The target variable was H_{post}. Due to the input variables value ranges varying by order of magnitude, the data was pre-processed using z-standardization, which is a numerical value that compares a specific input values relationship to the entire groups mean in terms of its standard deviation from the group mean. This reduces the Euclidean distance between data and improves the model training.

The target variable of GP is the ratio of H_{post}/H_{start}, representing the wave attenuation as a percentage of the initial wave height, before the vegetation patch. Ensuring that the GP does not predict the $H_{post}/H_{in} \geq 1$ means the second axiom is satisfied.

A logit transformation was applied to the target variable. The logit function is given by

$$\text{Logit}(p) = \ln\frac{p}{1-p} \quad \text{for } p \in (0,1) \tag{15}$$

where p is H_{post}/H_{start}. The logit transformation is a mathematical function that maps values from the real number line $(-\infty, +\infty)$ to a bounded interval (0, 1). Thus, the GP model is trained, and predictions are made in the unbounded domain. A desirable by-product of this transformation is the normalisation of the target variable, which improves the GP predictions.

The GpyTorch package (Gardner et al., 2018) was used to train the GP, the RBF kernel provided the best fit, using CASE1 (Table 1) as a calibration case. The RBF kernel is defined by:

$$\kappa(\mathbf{x}', \mathbf{x}'', I) = \exp\left(-\frac{1}{2}\sqrt{\left(\frac{x_1' - x_2''}{l_1}\right) + \cdots + \frac{x_n' - x_n''}{l_n}}\right) \tag{16}$$

where, this is essentially a scaled Euclidean norm with the hyperparameter vector $l = (l_1, \ldots, l_n)^\top$ is the vector of lengthscales for each dimension of the input parameter space. The lengthscale works as a scaling factor that can have the impact of erasing a particular dimension k from the equation above when the associated lengthscale $l_k \to \infty$.

As discussed before, the model hyperparameters must be tuned as they govern the fit of the GP. GpyTorch uses a gradient-descent based optimiser called Adam for this purpose (Kingma and Ba, 2014). Additionally, Adam optimiser has two parameters which can be calibrated: the learning rate and the number of iterations set at optimum values for this dataset of 0.003 and 2000, respectively, through simple search mechanisms, e.g. grid search.

In the case of AL sampling, the CMA-ES algorithm was explored in this study. The training sample size increased incrementally until the budget was reached. The mean standardised log loss (MSLL) and R^2 were calculated to assess the GP prediction capability.

Predicted values given by the GP model are given as a Gaussian distribution within the unbounded Logit transformed form. For better interoperability, the output must be transformed back to it's original format. This is achieved using the sigmoid function transformation, expressed mathematically by

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} \tag{17}$$

The PDF may be non-normal skewed once the target data has been transformed. We use the logit-normal density PDF derived by Atchison and Shen (1980) to address this.

$$f_X(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{x(1-x)}e^{-\frac{(\text{Logit}(x)-\mu)^2}{2\sigma^2}} \tag{18}$$

where μ and σ are the mean and standard deviation. We then use a quantile function to find the percentage probability function, giving the 50th and 95th percentiles. The median is also calculated.

An example of this is given in Fig. 6. Lastly, Fig. 7 shows a flowchart for the workflow schedule.

4. Results

4.1. Results of sampling methods, GP model training and predictive capability

In this section, we compare the three sampling methods used and their influence on the predictive capacity of the GP model. The highest-performing model is then selected.

Fig. 8 displays the GP model's skill scores (R^2 and MSLL) for the three sampling methods when using the synthetic dataset for testing. To statistically determine which method was most effective, we have adopted a standard statistical testing method, a one-sided Wilcoxon signed-rank test. Starting with the null hypothesis and assuming no difference between the three sampling methods, with the alternative to the hypothesis discerning which method was better. A Bonferroni correction (Abdi, 2007) was applied to adjust for multiple comparisons. The value P is the probability that the given null hypothesis is true. The value alpha is the threshold value used where $\alpha = 0.05/3 = 0.0167$.

The results revealed that for R^2, no statistically significant difference was observed between the three sampling methods. For the MSLL, the

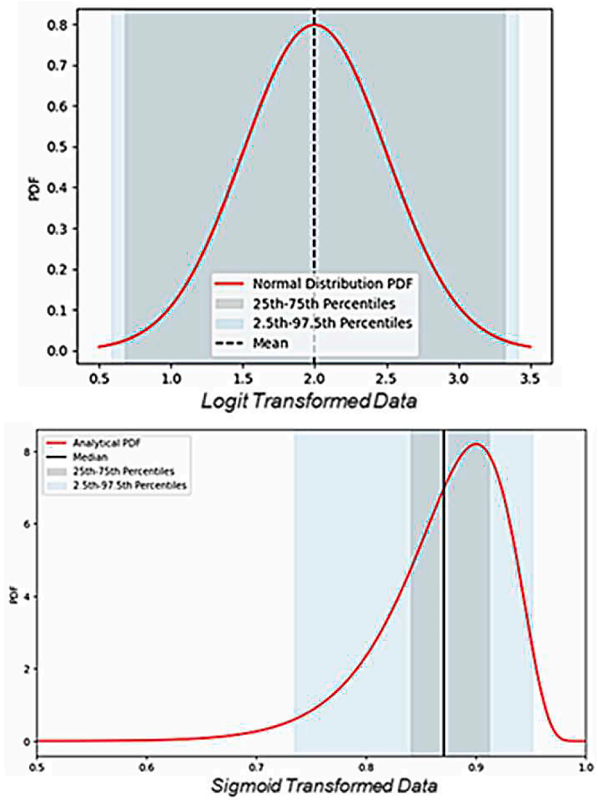

Fig. 6. Using example data with μ=2 and $\sigma = 0.5$ Top panel - Example of the probability density function of the target output from Logit transformed domain [-∞,∞]. Bottom panel – Example of the probability density function of target output with Sigmoid transformation domain applied to the Logit transformed data [0,1].

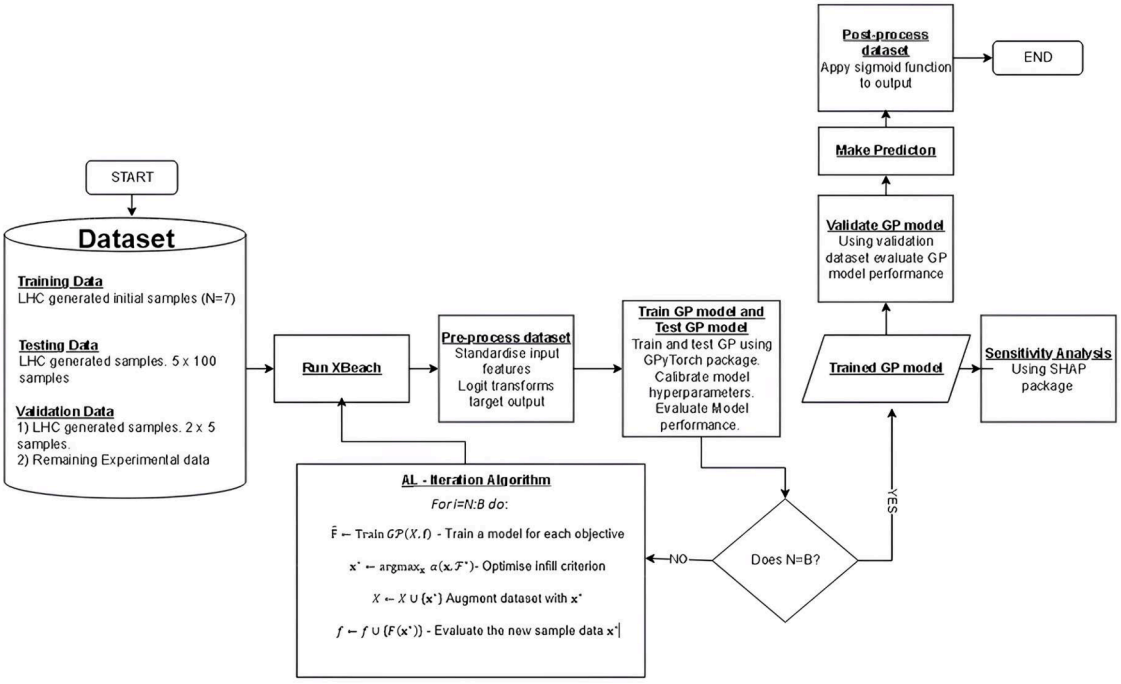

Fig. 7. Flow chart of complete project for the GP model training, testing and validation.

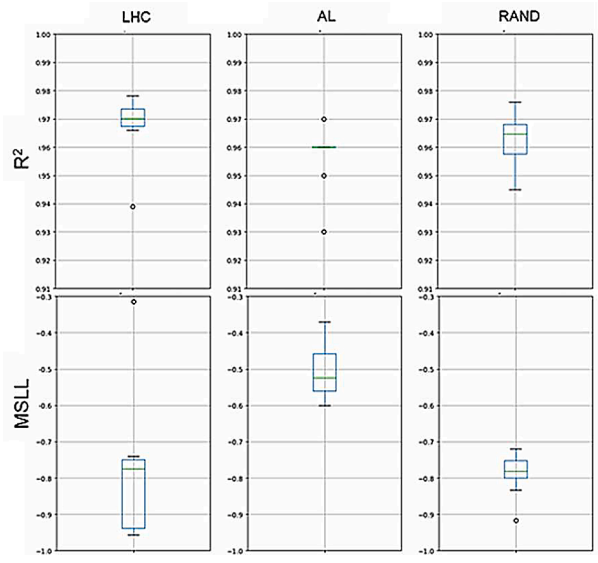

Fig. 8. Summary statistic boxplots for LHC, AL and RAND sampling LHC and XBeachX generated validation data. Top row – R^2, Bottom row – MSLL.

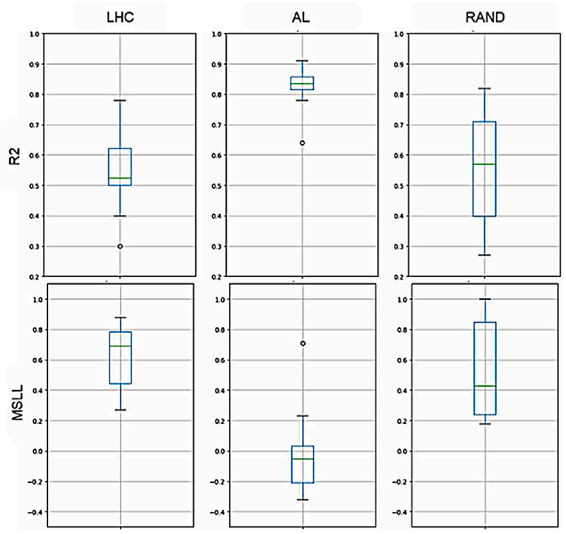

Fig. 9. Summary statistic boxplots for LHC, AL and RAND sampling for experimental validation data. Top row – R^2, Bottom row – MSLL.

test revealed that LHC and RAND were statistically significantly less than AL sampling ($p < \alpha$) whilst there was no statistically significant difference between RAND and LHC sampling ($p < \alpha$).

Previous studies have found the performance of AL sampling to outperform other sampling methods (Hansen and Ostermeier, 2001; Hansen and Kern, 2004; Hansen, 2009), but we did not observe this in the synthetic dataset. This may be because of two reasons: the underlying function is easy to learn with this number of data points, and the synthetically generated test set is (almost) uniformly distributed in the input space due to the nature of LHC.

However, comparing the three GP models learned using distinct sampling methods to predict the experimental validation data demonstrates that AL is superior in this scenario. Fig. 9 displays summary statistics boxplots of each sampling method and its skill scores for predicting the validation experimental data. The R2 and MSLL of the AL is significantly improved compared with RAND and LHC sampling, which suggests the AL method better generalises to unseen datasets. The results of the Wilcoxon test further support this. The results revealed for R^2 AL was statistically significantly greater than LHC and AL sampling ($p < \alpha$). For the MSLL, the test revealed that AL was statistically significantly less than LHC and RAND sampling ($p < \alpha$) whilst there was no statistically significant difference between LHC and RAND sampling ($p > \alpha$).

A model trained on a set that covers the input space efficiently can identify regions of greatest changes in the output function responses with the minimal number of data points is likely to perform better than other alternative training schemes. In other words, the training data distribution directly impacts the effectiveness of the models. LHC sampling, designed for systematically exploring input spaces, performs well

when testing data aligns with training data. However, LHC is agnostic of the function responses. So, the designs generated may not be efficient in identifying regions with the greatest change in function values. Consequently, LHC sampling underperforms for the experimental dataset since the input samples distribution is almost uniform across the input space.

In contrast, a uniform random sampling method may fail to generate a good coverage of the space when the number of samples is low across the domain. Because of this high discrepancy random design, it is usually expected to diverge from LHC's structured representation and the specific distribution of experimental data. The success of AL sampling depends on adapting to sampled data patterns. This means it would explore areas where functions change more drastically than areas where

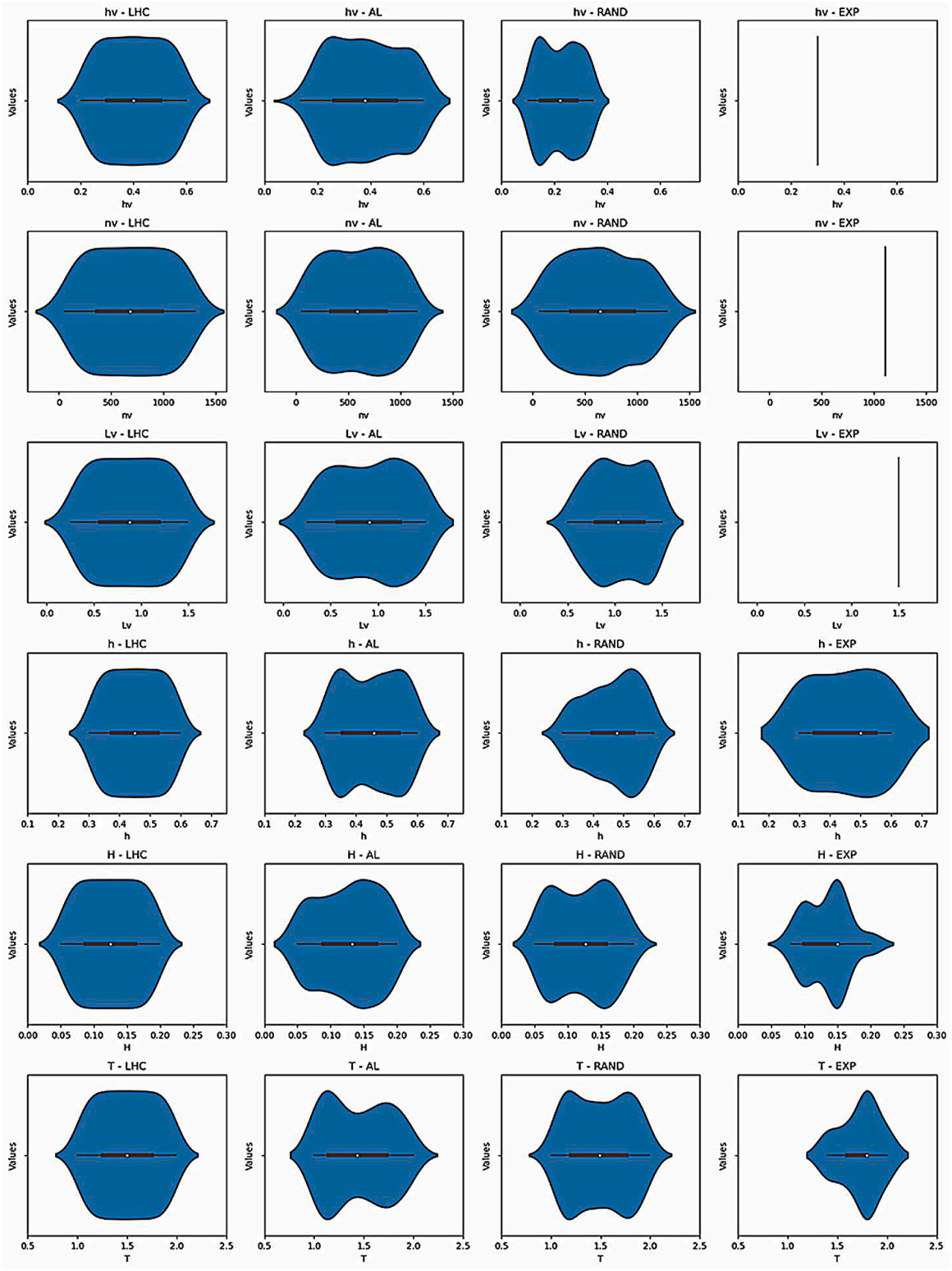

Fig. 10. Violin Plots displaying each input variables distribution for the three sampling strategies, LHC, AL and Random. The labelling of each subplot has the convention A-B, where A is the input parameter name used in XBeachX simulation, and B is the sampling strategy.

the response is flat. Therefore, models trained on LHC should better represent the true underlying function for the given number of data points. Nonetheless, the choice of sampling method should be empirically assessed based on the dataset's characteristics and research objectives.

Fig. 10 supports the hypothesis above. The violin plots demonstrate that clustering occurs in the AL dataset for all the input features of the ten cases tested. Whilst this is also observed in the random sampling dataset, this clustering is inherently random and often does not provide samples to the area of highest uncertainty.

The Gaussian Process with Active Learning data sampling method (GP-AL) generalises best to both XBeachX and experimental datasets; the remainder of the study focuses solely on the GP-AL approach. Table 2 highlights the impact of the AL process for each of the ten cases used to train the GP model. In all cases, there is a significant increase in both R^2 and MSLL. This is further demonstrated graphically in Fig. 11.

4.2. Making predictions using GP-AL model

The trained and tested GP-AL model performance is compared with the experimental data (Table 1), specifically the GP test data. Fig 12 compares the predictions of the GP-AL model (top) and XBeachX model (bottom) with the experimental data, not used for XBeachX calibration. Although the GP-AL model cannot outperform XBeachX, the GP-AL model scores equally, with R^2=0.84 observed for both models. Furthermore, all GP model predictions are within one standard deviation of the actual values, and the scattering of data closely resembles the XBeachX model predictions. This suggests, within the bounds of training and testing data and the scope of the trained GP model, that the accuracy of the GP model is comparable to a process-based model for the prediction of wave attenuation through rigid vegetation.

While the GP-AL model was trained and tested for point predictions, we now demonstrate that the model can successfully predict wave attenuation along the length of the vegetation patch. One significant benefit of numerical models is offering such spatial modelling capabilities. It is, therefore, desirable for ML models to offer similar spatial analysis. Including L_v as an input enables cross-shore predictions with the GP-AL model. The trained GP-AL model was used to predict the cross-shore wave attenuation for two unseen cases, given in Table 3.

Notably, all conditions remained constant, with only L_v varying for each case. The results are shown in Fig. 13, which displays both the logit-transformed and sigmoid-transformed output values. The GP-AL model correctly predicts increasing wave attenuation as the wave propagates further into the vegetation patch. All predictions lay within either the 90th or 95th percentile. The PDF for each prediction point in TEST02 is found in Fig. 14, highlighting the skewed normal distribution of the sigmoid-transformed data.

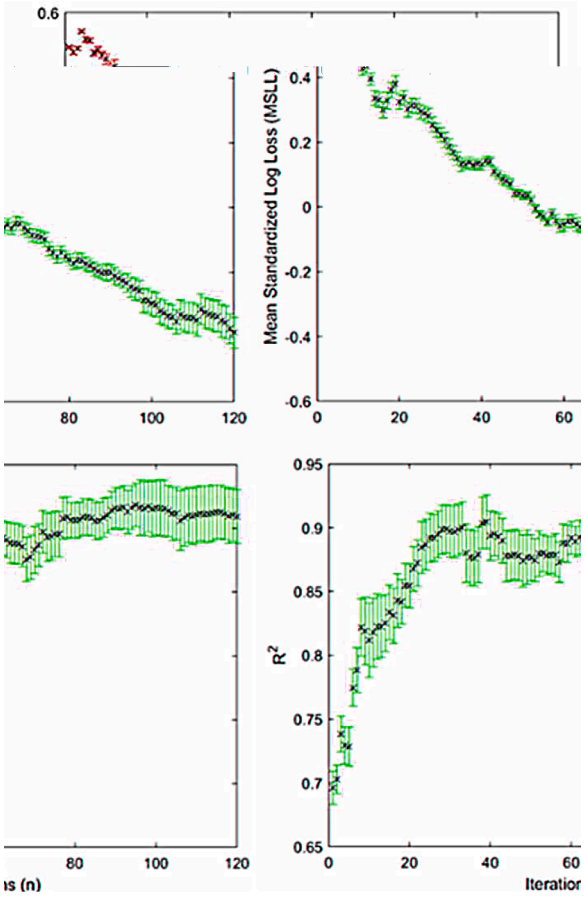

Fig. 11. For CASE10 the mean values and standard deviation of MSLL (top panel) and R^2 (tbottom panel) for each test case over each AL iteration. Incremental improvement was observed.

4.3. Sensitivity analysis

The results shown in Section 4.2 confirm that the GP-AL can yield accurate predictions of wave attenuation on coastal vegetation. However, it can be useful to explore how the GP-AL model ranks the contribution of each input parameter, ranks parameter interaction, and impacts the output variable. This understanding can be achieved

Table 2
Skill scores for all 10 cases of GP model performance for the AL process. The Pre AL scores and Post AL scores are displayed. Lower values of MSLL indicate lower uncertainty and higher R^2 indicates higher correlation between predicted and actual values.

	Pre AL LHC i = 7				Post AL $i = 127$			
	XBeachX		EXP		XBeachX		EXP	
CASE	R^2	MSLL	R^2	MSLL	R^2	MSLL	R^2	MSLL
CASE1	0.07	4.76	−3.30	8.30	0.96	−0.37	0.64	0.71
CASE2	0.28	0.80	0.52	0.51	0.96	−0.55	0.83	0.23
CASE3	0.70	0.33	0.28	0.70	0.97	−0.56	0.81	−0.01
CASE4	0.74	0.30	−0.09	0.93	0.95	−0.39	0.78	0.03
CASE5	0.56	0.76	−2.70	2.40	0.96	−0.45	0.85	−0.15
CASE6	0.68	0.34	0.57	0.38	0.96	−0.56	0.90	−0.26
CASE7	0.26	1.14	0.44	1.11	0.96	−0.50	0.84	−0.09
CASE8	0.67	0.35	0.24	0.79	0.96	−0.60	0.91	−0.32
CASE9	0.71	0.44	−0.12	0.93	0.96	−0.60	0.86	−0.23
CASE10	0.80	0.47	0.32	0.72	0.93	−0.48	0.83	0.03
Average	**0.55**	**0.97**	**−0.38**	**1.68**	**0.96**	**−0.51**	**0.83**	**−0.01**
Lower	**0.07**	**0.30**	**−3.30**	**0.38**	**0.93**	**−0.60**	**0.64**	**−0.32**
Upper	**0.80**	**4.76**	**0.57**	**8.30**	**0.97**	**−0.37**	**0.91**	**0.71**

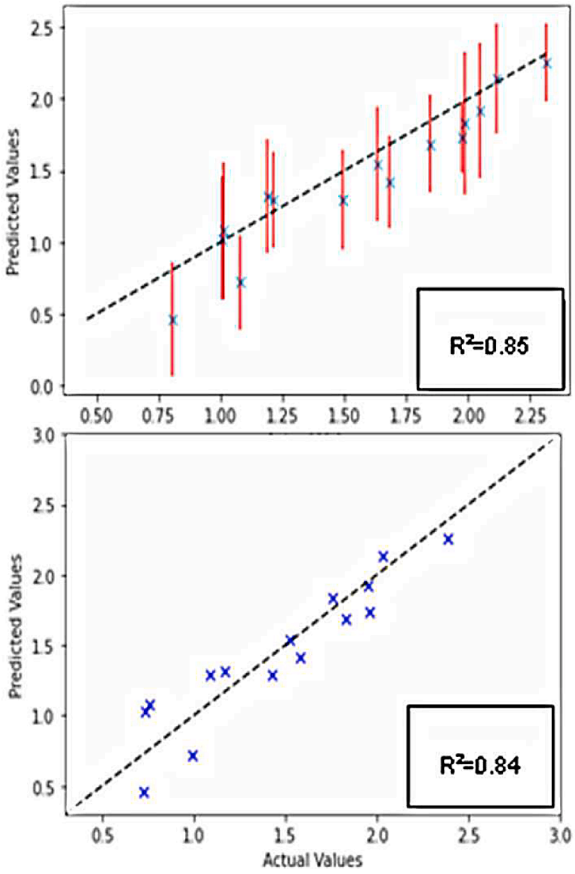

Fig. 12. Predicted vs Actual values (Experimental data) with logit transformation applied. Top panel – GP-AL model predictions. Blue crosses in the left figure indicate the mean prediction of the GP, and the red error bars depict uncertainty (± one standard deviation). Bottom panel – XBeachX model predictions.

Table 3
Input parameters for cross-shore predictions using the GP-AL model. Conditions were randomly generated.

Case	h_v [m]	Nv [stems/m2]	Lv [m]	h [m]	H_0 [m]	T [s]
Test01	0.29	1104	0.5 – 1.5	0.34	0.10	1.8
Test02	0.182	900	0.5	0.58	0.096	1.5

through a sensitivity analysis of each input parameter's impact on the output variable. Below are the global and local sensitivity analysis results for the GP-AL model. The outcome and interpretation of the sensitivity analysis are compared with existing knowledge from previous literature.

Fig. 15 illustrates the global sensitivity analysis results using the SHAP package. While the GP-AL model was trained on a small sample size, the SHAP analysis was performed on 1000 test cases generated by the trained GP-AL model. Each input variable is displayed in descending order of influence on wave attenuation, with the x-axis indicating the SHAP values that represent each feature's impact on wave attenuation. The SHAP values indicate the impact a variable has on the prediction. Negative SHAP values represent a lower-than-average prediction, and positive SHAP values represent higher-than-average predictions of wave attenuation. The legend on the right-hand side shows the colour codes for each input feature value. Blue represents input variables much lower than the average for that variable. Red represents higher-than-average values for that variable. For example, for smaller values of h (blue), there is a larger wave attenuation than the average wave attenuation observed. Likewise, For larger values of h (red) there is less wave attenuation than the average wave attenuation.

The analysis reveals that the total water depth h has the most significant influence on wave attenuation. The variable is positively correlated to wave attenuation, with lower values of h resulting in more significant wave attenuation. Contrastingly, all other variables negatively correlate to wave attenuation, with higher values reducing wave attenuation. It is worth noting that Lv is the least influential variable on wave attenuation compared to the other variables. However, it should be noted that this could be due to the limited range of L_v used in this study (0.5 m – 1.5 m).

A more detailed analysis of each feature's influence over wave attenuation is provided in Fig. 16, where SHAP dependence plots are given. The left-side y-axis displays the SHAP value, and the x-axis displays the standardised value of the considered input feature. For Fig. 16 (A)–(F), the right-side y-axis shows the most interacted feature per the colour-coded legend. Fig. 16(G) presents the correlation between h and h_v. These dependence pots offer a more detailed look at each input feature's impact on the wave attenuation. The OLS coefficient are displayed in the figure, which provide a quantitative method for comparison.

In the present study, larger incident wave height (H_{start}) values result in more significant wave attenuation, which is consistent with the results of several other studies (Anderson and Smith, 2014; Möller et al., 2014; Tschirky et al., 2000). However, van Wesenbeeck et al. (2022) reported that incident wave height had no impact on the amount of wave attenuation. The impact of wave period on wave attenuation is inconclusive. In this study, we find a negative correlation between wave period and wave attenuation, a result also found by Koftis et al. (2013) who also found that longer wavelengths resulted in greater wave attenuation. However, other studies have reported no notable correlation between wave period and wave attenuation (Tschirky et al., 2000; Möller et al., 1999). The most influential hydrodynamic parameter, total water depth (h), shows a highly positive correlation with wave attenuation. This is expected as smaller water depths have increased wave attenuation in several other studies (Tschirky et al., 2001; van Wesenbeeck et al., 2022; Möller et al., 2014; Yang et al., 2012). However, it should be noted that considering water depth in isolation can be misleading. Multiple studies have shown that vegetation height to water depth ratio (submergence ratio) is a key driver in wave attenuation, with emergent vegetation conditions producing larger wave attenuation (e.g. Ysebaert et al., 2011; Möller et al., 2014; Anderson and Smith, 2014).

To explore this further, in Fig. 16(G) the correlation between h_v and h is given, which displays a negative correlation to wave attenuation. There is a significant scattering of h values, with no discernible trend observed. This suggests that the GP model does not capture the relationship between submergence ratio and wave attenuation. Fig. 16(B) demonstrates that L_v is the most closely interacted parameter with h_v, which is not expected. The submergence ratio of hv/h is a widely known as a governing parameter of wave attenuation. As shown in Fig. 15, L_v has the least impact on wave attenuation. This observation is reinforced by the results shown in Fig. 16(C), which indicates a weakly negative correlation with wave attenuation. Due to the limited vegetation patch length used in this study, the effects of Lv may not be adequately captured in our modelling. Lastly, N_v negatively correlates with wave attenuation and interacts strongly with L_v. This agrees with other studies which concluded that increasing N_v increased wave attenuation (Anderson and Smith, 2014; Tschirky et al., 2000).

The preceding discussion focused on the global interoperability of the GP model. Below, local interoperability is considered. Fig. 17 demonstrates the local interoperability of the GP model using Case 1 from Table 1. The plot reveals the individual impact of each input feature on wave attenuation predictions. The output value for the particular case is displayed in bold font, and the base value represents the mean prediction from all cases modelled. Each input feature either contributes negatively or positively to the output value.

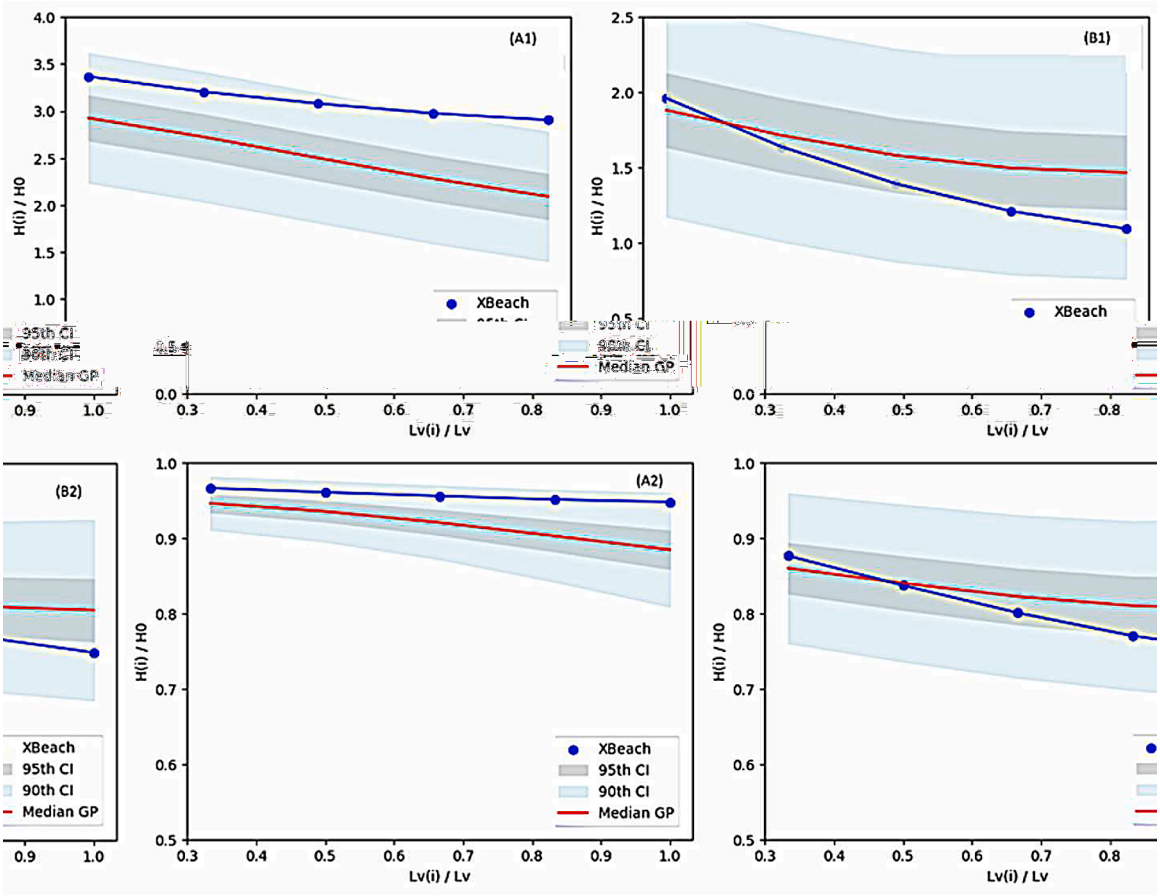

Fig. 13. Cross-shore predictions of wave attenuation for two unobserved datasets, which were not included in model training and or testing stages. The blue line is the XBeachX output, the red lines are GP model outputs, and the 25th-75th percentile are shaded grey and the 2.5th-97.5th are shaded light blue. The top figure represents prediction in with logit transformation applied. The bottom figures represent post-processed data, with sigmoid transformed applied. Fig. 12(A) represents TEST02 and Figure (B) represents TEST02 in Table 3.

Fig. 17 is displaying the local interoperability of a specific case, which is used for analysing individual cases and making inferences regarding the predicted output value. This can aid in calibrating ML models and selecting appropriate variables. Additionally, the plot demonstrates that while the global interoperability plots suggest h and L_v are the most and least impactful features, in this specific case, h_v is the most impactful on the prediction. In Fig. 17 H_{start} and h both positively contribute and h_v, n_v, L_v and T negatively contribute to wave attenuation. In this scenario, h_v contributes the most to reducing the output value, and H_{start} contributes the least, highlighting the benefit of analysing cases locally.

5. Conclusion

This study presents a framework of a data-agnostic GP modelling approach by applying a GP model to numerical model data to predict wave attenuation on rigid, submerged and emergent vegetation and wave attenuation results. We aimed to address four issues in this study as stated in Section 1. These have been addressed as follows:

(i) The GP model could predict wave attenuation for a range of wave conditions, representative of conditions found in estuaries and other sheltered wave environments, to within 90% confidence. The prediction time of the trained GP model is significantly less than that of XBeachX, with predictions being determined in a matter of seconds. The testing and validation of the GP model performance against unseen XBeachX highlights that the model is comparable to the XBeachX, for the range of conditions explored in this study.

(ii) The uncertainty of the GP predictions has been provided This can provide essential design thresholds for practitioners and engineers.

(iii) The concepts and methods applied in this study can be considered a successful proof of concept for applying GP models to other regression-type problems in coastal engineering. Providing a data-agnostic, methodological framework that can be applied to a broader range of coastal engineering uses. The benefits of employing an AL sampling method to optimise the training of a GP model were also highlighted. This limits the number of experimental or numerical modelling cases that are required for model training. Care must be taken to ensure the data and problems are suitable for GP approaches.

(iv) An understanding of each input feature's impacts on the GP model were deduced through a global and local sensitivity analysis. This interoperability of the ML compliments the high predictive capability.

Future work is needed to extend the applicability of the current GP-AL model. In particular, the model assumed a constant drag coefficient, making it unsuitable for environmental conditions with significantly different vegetation, hydrodynamic conditions, and drag coefficients. Additionally, the effects of plant rigidity are not explored as the study was based on rigid vegetation with a constant flexural rigidity. However, the rigidity of many rigid saltmarsh and mangrove plants does not significantly deviate from the selected plant rigidity. The study is also

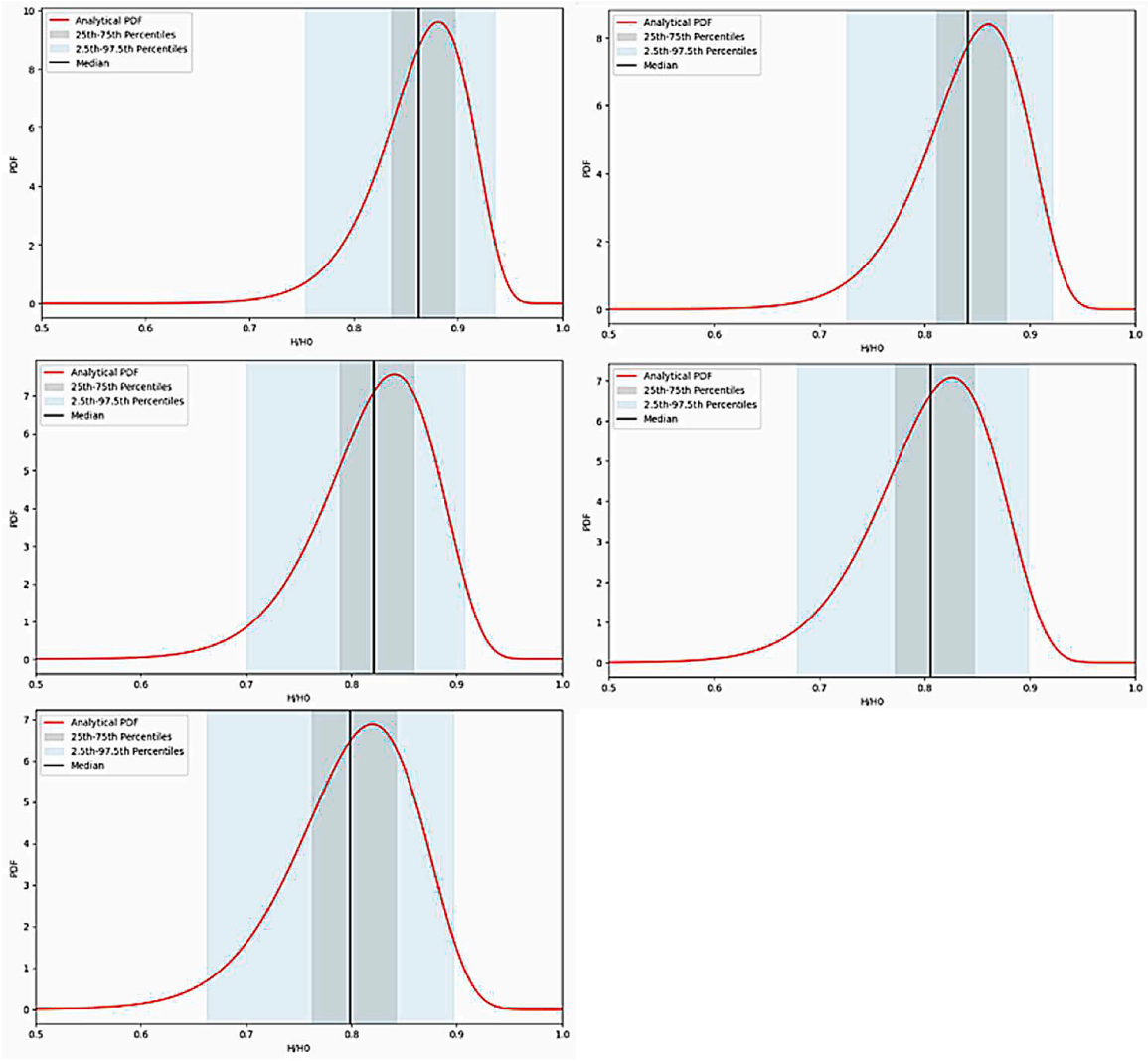

Fig. 14. PDF function is plotted in red for the sigmoid transformed predicted output for the case TEST02 in Table 3. The 25th-75th percentiles are shaded grey and the 2.5th-97.5th percentiles are shaded light blue, which have been calculated through numerical integration.

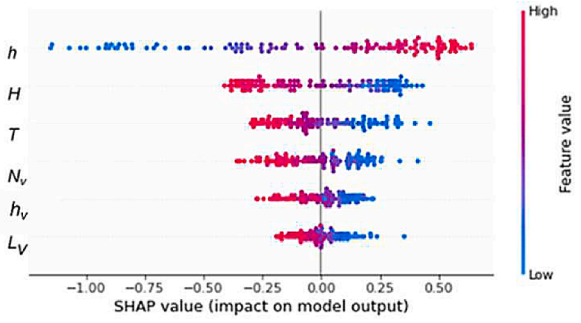

Fig. 15. SHAP Variance importance plot. SHAP value on X-axis indicates a positive or negative effect on the target variable. The left side y-axis shows the features. The right side y-axis indicates if the feature value was high or low compared to the average. Features values shown are standardised.

limited to regular waves where randomness of the incident waves can play a part in wave attenuation on vegetation. GP modelling methods should be further assessed for feasibility, particularly since the approach is unsuitable for non-stationary and non-continuous datasets.

Authors statement

No tools were used in the construction of this manuscript.

CRediT authorship contribution statement

Kristian Ions: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Alma Rahat:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Dominic E. Reeve:** . **Harshinie Karunarathna:** Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Kristian Ions reports financial support was provided by Swansea University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

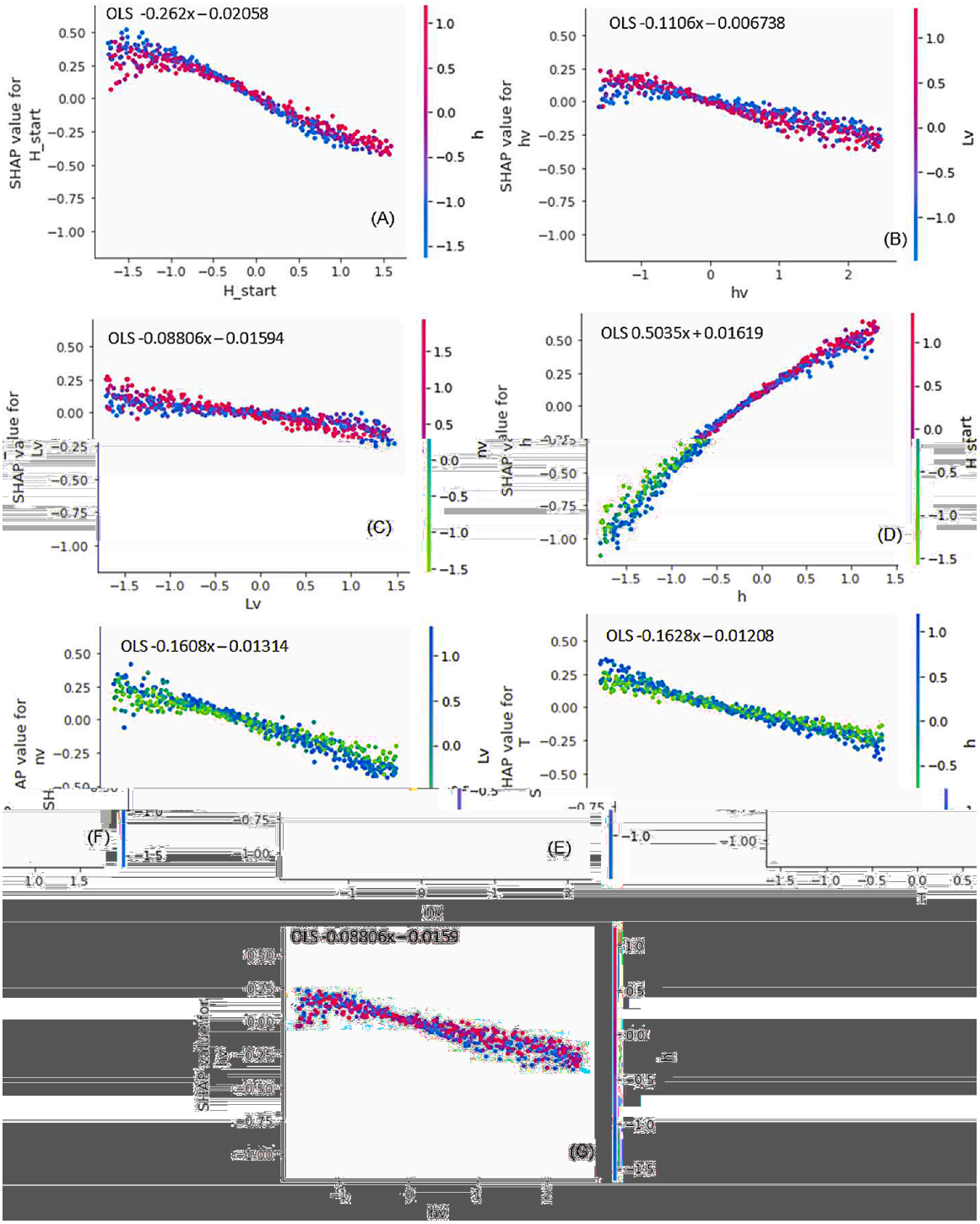

Fig. 16. SHAP Dependence plot. X-axis gives the feature considered in the analysis. Left side y-axis gives the SHAP value for the feature's corresponding value, which is the variance of the value from its mean value. Right side y-axis gives the most interacted feature of the x-axis and the corresponding value. The OLS coefficients are given above each plot. Features values shown are standardised. (G) represents SHAP Dependence plot of h_v vs h.

Fig. 17. SHAP Individual value plot for case 1 in Table 1. The Feature values shown are standardised.

Data availability

Data will be made available on request.

Acknowledgments

KI's PhD is supported by the Engineering and Physical Sciences Research Council (EPSRC) UK Doctoral Training Partnership of Swansea

University. The authors also thank Dr Thomas van Veelen of Twente University, The Netherlands, for providing the experimental data used in this study, without which this study would not have been /possible.

References

Abdi, H., 2007. Bonferroni and Šidák corrections for multiple comparisons. Encycl. Meas. Stat. 3 (01), 2007.

Abdolali, A., Roland, A., Van Der Westhuysen, A., Meixner, J., Chawla, A., Hesser, T., Smith, J.M., Sikiric, M.D., 2020. Large-scale hurricane modeling using domain decomposition parallelization and implicit scheme implemented in WAVEWATCH III wave model. Coastal Eng. 157, 103656 https://doi.org/10.1016/j.coastaleng.2020.103656.

Anderson, M.E., Smith, J.M., 2014. Wave attenuation by flexible, idealised salt marsh vegetation. Coastal Eng. 83, 82–92.

Anderson, M.E., Smith, J.M., & McKay, S.K. (2011). Wave dissipation by vegetation.

Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Paz, D., Invariant Risk Minimization (2019).

Atchison, J., Shen, S.M., 1980. Logistic-normal distributions: Some properties and uses. Biometrika 67 (2), 261–272.

Augustin, L.N., Irish, J.L., Lynett, P., 2009. Laboratory and numerical studies of wave damping by emergent and near-emergent wetland vegetation. Coastal Eng. 56 (3), 332–340.

Baldock, T.E., Holmes, P., Bunker, S., Van Weert, P., 1998. Cross-shore hydrodynamics within an unsaturated surf zone. Coast. Eng. 34 (3–4), 173–196. https://doi.org/10.1016/S0378-3839(98)00017-9.

Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C., Silliman, B.R., 2011. The value of estuarine and coastal ecosystem services. Ecol. Monogr. 81 (2), 169–193. https://doi.org/10.1890/10-1510.1.

Bennett, W.G., Horrillo-Caraballo, J.M., Fairchild, T.P., van Veelen, T.J., Karunarathna, H., 2023. Saltmarsh vegetation alters tidal hydrodynamics of small estuaries. Appl. Ocean Res. 138, 103678.

Bennett, W.G., van Veelen, T.J., Fairchild, T.P., Griffin, J.N., Karunarathna, H., 2020. Computational modelling of the impacts of saltmarsh management interventions on hydrodynamics of a small macro-tidal estuary. J. Mar. Sci. Eng. 8 (5), 373.

Beuzen, T., Goldstein, E.B., Splinter, K.D., 2019. Ensemble models from machine learning: an example of wave runup and coastal dune erosion. Nat. Hazards Earth Syst. Sci. 19 (10), 2295–2309.

Camfield, F.E., 1983. Wind-wave growth with high friction. J. Waterway, Port, Coastal, Ocean Eng. 109 (1), 115–117.

Chau, K., 2006. A review on the integration of artificial intelligence into coastal modeling. J. Environ. Manag. 80 (1), 47–57.

Cohen-Shacham, E., Walters, G., Janzen, C., Maginnis, S., 2016. Nature-based solutions to address global societal challenges. IUCN: Gland, Switzerland 97, 2016–2036.

Dalrymple, R.A., Kirby, J.T., Hwang, P.A., 1984. Wave diffraction due to areas of energy dissipation. J. Waterway, Port, Coastal, Ocean Eng. 110 (1), 67–79.

Dean, R.G., Dalrymple, R.A., 1991. *Water Wave Mechanics for Engineers and Scientists* (Vol. 2). World Scientific Publishing Company.

den Bieman, J.P., van Gent, M.R., van den Boogaard, H.F, 2021. Wave overtopping predictions using an advanced machine learning technique. Coastal Eng. 166, 103830.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Duvenaud, D. (2014). Automatic model construction with Gaussian processes (Doctoral dissertation).

Dwarakish, G.S., Nithyapriya, B., 2016. Application of soft computing techniques in coastal study–A review. J. Ocean Eng. Sci. 1 (4), 247–255.

Fairchild, T.P., Bennett, W.G., Smith, G., Day, B., Skov, M.W., Möller, I., Griffin, J.N., 2021. Coastal wetlands mitigate storm flooding and associated costs in estuaries. Environ. Res. Lett. 16 (7), 074034.

Gardner, J.R., Pleiss, G., Bindel, D., Weinberger, K.Q., Wilson, A.G., 2018. GPyTorch: blackbox matrix-matrix Gaussian process inference with GPU acceleration. Adv. Neural Inf. Process. Syst.

Ghisalberti, M., Nepf, H.M., 2002. Mixing layers and coherent structures in vegetated aquatic flows. J. Geophys. Res. Oceans 107 (C2), 3-1.

Giri, C., Ochieng, E., Tieszen, L.L., Zhu, Z., Singh, A., Loveland, T., Duke, N., 2011. Status and distribution of mangrove forests of the world using earth observation satellite data. Global Ecol. Biogeogr. 20 (1), 154–159.

Goldstein, E.B., Coco, G., Plant, N.G., 2019. A review of machine learning applications to coastal sediment transport and morphodynamics. Earth-Sci. Rev. 194, 97–108.

Gracia, S., Olivito, J., Resano, J., Martin-del-Brio, B., Alfonso, M., Alvarez, E., 2021. Improving accuracy on wave height estimation through machine learning techniques. Ocean Eng. 236, 108699 https://doi.org/10.1016/j.oceaneng.2021.108699.

Hansen, 2009. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, pp. 2389–2396.

Hansen, Kern, 2004. Evaluating the CMA evolution strategy on multimodal test functions. In: Eighth International Conference on Parallel Problem Solving from Nature PPSN VIII, Proceedings, pp. 282–291.

Hansen, N., Ostermeier, A., 2001. Completely derandomized self-adaptation in evolution strategies. Evol. Comput. 9 (2), 159–195.

Himes-Cornell, A., Pendleton, L., Atiyah, P., 2018. Valuing ecosystem services from blue forests: a systematic review of the valuation of salt marshes, sea grass beds and mangrove forests. Ecosyst. Serv. 30, 36–48.

Holthuijsen, L.H., Booij, N., Herbers, T.H.C., 1989. A prediction model for stationary, short-crested waves in shallow water with ambient currents. Coastal Eng. 13 (1), 23–54. https://doi.org/10.1016/0378-3839(89)90031-8. URL: http://linkinghub.elsevier.com/retrieve/pii/0378383989900318.

Hosseinzadeh, S., Etemad-Shahidi, A., Koosheh, A., 2021. Prediction of mean wave overtopping at simple sloped breakwaters using kernel-based methods. J. Hydroinf. 23 (5), 1030–1049.

Hsieh, W.W., 2009. Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels. Cambridge university press. https://doi.org/10.1016/S0029-8018(97)10025-7.

Hu, Z., Suzuki, T., Zitman, T., Uittewaal, W., Stive, M., 2014. Laboratory study on wave dissipation by vegetation in combined current–wave flow. Coastal Eng. 88, 131–142.

IPCC, 2021: Climate Change 2021: The Physical Science Basis. Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. (eds), Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

Jadhav, R.S., Chen, Q., Smith, J.M., 2013. Spectral distribution of wave energy dissipation by salt marsh vegetation. Coast. Eng. 77, 99–107. https://doi.org/10.1016/j.coastaleng.2013.02.013.

James, S.C., Zhang, Y., O'Donncha, F, 2018. A machine learning framework to forecast wave conditions. Coastal Eng. 137, 1–10. https://doi.org/10.1016/j.coastaleng.2018.03.004.

Kathiresan, K., Rajendran, N., 2005. Coastal mangrove forests mitigated tsunami. Estuar Coast Shelf Sci. 65 (3), 601–606. https://doi.org/10.1016/j.ecss.2005.06.0222.

Keulegan, G.H., Carpenter, L.H., 1956. Forces on Cylinders and Plates in an Oscillating Fluid. National Bureau of Standards.

Kim, T., Kwon, Y., Lee, J., Lee, E., Kwon, S., 2022. Wave attenuation prediction of artificial coral reef using machine-learning integrated with hydraulic experiment. Ocean Eng. 248, 110324.

Kingma, D.P., & Ba, J. (2014). Adam: a method for stochastic optimisation. *arXiv preprint arXiv:1412.6980*.

Kirwan, M.L., Walters, D.C., Reay, W.G., Carr, J.A., 2016. Sea level driven marsh expansion in a coupled model of marsh erosion and migration. Geophys. Res. Lett. 43 (9), 4366–4373.

Kobayashi, N., Raichle, A.W., Asano, T., 1993. Wave attenuation by vegetation. J. Waterway, Port, Coastal Ocean Eng. 119 (1), 30–48.

Koftis, T., Prinos, P., Stratigaki, V., 2013. Wave damping over artificial Posidonia oceanica meadow: a large-scale experimental study. Coastal Eng. 73, 71–83.

Li, C.W., Yan, K., 2007. Numerical investigation of wave– current–vegetation interaction. J. Hydraul. Eng. 133 (7), 794–803.

Losada, I.J., Maza, M., Lara, J.L., 2016. A new formulation for vegetation-induced damping under combined waves and currents. Coastal Eng. 107, 1–13.

Luhar, M., Infantes, E., Nepf, H., 2017. Seagrass blade motion under waves and its impact on wave decay. J. Geophys. Res. Oceans 122 (5), 3736–3752.

Luhar, M., Nepf, H.M., 2016. Wave-induced dynamics of flexible blades. J. Fluids Struct. 61, 20–41.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.

Maji, S., Senapati, A., Mondal, A., 2022. Investigation and validation of flow characteristics through emergent vegetation patch using machine learning technique. Evolution in Computational Intelligence. Springer, Singapore, pp. 131–139.

Maza, B., Rodes-Blanco, M., Rojas, E., 2022. Aboveground biomass along an elevation gradient in an evergreen Andean–Amazonian Forest in Ecuador. Front. For. Global Change 5, 738585.

Maza, M., Lara, J.L., Losada, I.J., 2013. A coupled model of submerged vegetation under oscillatory flow using Navier–Stokes equations. Coastal Eng. 80, 16–34.

Maza, M., Lara, J.L., Losada, I.J., Ondiviela, B., Trinogga, J., Bouma, T.J., 2015. Large-scale 3-D experiments of wave and current interaction with real vegetation. Part 2: experimental analysis. Coastal Eng. 106, 73–86.

McCall, R.T., Masselink, G., Poate, T.G., Roelvink, J.A., Almeida, L.P., 2015. Modelling the morphodynamics of gravel beaches during storms with XBeach-G. Coastal Eng. 103, 52–66.

McCall, R.T., Masselink, G., Poate, T.G., Roelvink, J.A., Almeida, L.P., Davidson, M., Russell, P.E., 2014. Modelling storm hydrodynamics on gravel beaches with XBeach-G. Coastal Eng. 91, 231–250.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245. https://doi.org/10.2307/1268522. ISSN 0040-1706. JSTOR 1268522. OSTI 5236110.

Mcowen, C.J., Weatherdon, L.V., Van Bochove, J.W., Sullivan, E., Blyth, S., Zockler, C., Fletcher, S., 2017. A global map of saltmarshes. Biodivers. Data J. (5).

Mendez, F.J., Losada, I.J., 2004. An empirical model to estimate the propagation of random breaking and nonbreaking waves over vegetation fields. Coastal Eng. 51 (2), 103–118.

Minuzzi, F.C., Farina, L., 2023. A deep learning approach to predict significant wave height using long short-term memory. Ocean Model. 181, 102151.

Möller, I., Kudella, M., Rupprecht, F., Spencer, T., Paul, M., Van Wesenbeeck, B.K., Schimmels, S., 2014. Wave attenuation over coastal salt marshes under storm surge conditions. Nat. Geosci. 7 (10), 727–731.

Möller, I., Spencer, T., French, J.R., Leggett, D.J., Dixon, M., 1999. Wave transformation over salt marshes: a field and numerical modelling study from North Norfolk, England. Estuar. Coast. Shelf Sci. 49 (3), 411–426.

Molnar, C., 2020. Interpretable Machine Learning. Lulu. com.

Morison, J.R., Johnson, J.W., Schaaf, S.A., 1950. The force exerted by surface waves on piles. J. Pet. Technol. 2 (05), 149–154.
Mork, M., 1996. Wave attenuation due to bottom vegetation. Waves and Nonlinear Processes in Hydrodynamics. Springer Netherlands, Dordrecht, pp. 371–382.
Nardin, W., Lera, S., Nienhuis, J., 2020. Effect of offshore waves and vegetation on the sediment budget in the Virginia Coast Reserve (VA). Earth Surf. Process. Landforms 45 (12), 3055–3068.
Nepf, H.M., 2012. Flow and transport in regions with aquatic vegetation. Annu. Rev. Fluid Mech. 44, 123–142.
Ozeren, Y., Wren, D.G., Wu, W., 2014. Experimental investigation of wave attenuation through model and live vegetation. J. Waterway, Port, Coastal Ocean Eng. 140 (5), 04014019.
Panchigar, D., Kar, K., Shukla, S., et al., 2022. Machine learning-based CFD simulations: a review, models, open threats, and future tactics. Neural Comput. Appl. 34, 21677–21700. https://doi.org/10.1007/s00521-022-07838-6.
Pontee, N., Narayan, S., Beck, M.W., Hosking, A.H., 2016. Nature-based solutions: lessons from around the world. In: Proceedings of the Institution of Civil Engineers-Maritime Engineering, 169. Thomas Telford Ltd, pp. 29–36.
Price, W.A., Tomlinson, K.W., Hunt, J.N., 1968. The effect of artificial seaweed in promoting the build-up of beaches. Coastal Eng. 570–578.
Quartel, S., Kroon, A., Augustinus, P.G.E.F., Van Santen, P., Tri, N.H., 2007. Wave attenuation in coastal mangroves in the Red River Delta, Vietnam. J. Asian Earth Sci. 29 (4), 576–584.
Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*.
Roelvink, D., Reniers, A., Van Dongeren, A.P., De Vries, J.V.T., McCall, R., Lescinski, J., 2009. Modelling storm impacts on beaches, dunes and barrier islands. Coastal Eng. 56 (11–12), 1133–1152.
Salehi, H., Burgueño, R., 2018. Emerging artificial intelligence methods in structural engineering. Eng. Struct. 171, 170–189.
Shapley, L.S., 1953. Stochastic games. Proc. Natl. Acad. Sci. 39 (10), 1095–1100.
Shepard, C.C., Crain, C.M., Beck, M.W., 2011. The protective role of coastal marshes: a systematic review and meta-analysis. PLoS ONE 6 (11), e27374. https://doi.org/10.1371/journal.pone.0027374.
Short, F., Carruthers, T., Dennison, W., Waycott, M., 2007. Global seagrass distribution and diversity: a bioregional model. J. Exp. Mar. Biol. Ecol. 350 (1–2), 3–20.
Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction. MIT press.
Sutton-Grier, A.E., Wowk, K., Bamford, H., 2015. Future of our coasts: the potential for natural and hybrid infrastructure to enhance the resilience of our coastal communities, economies and ecosystems. Environ. Sci. Policy 51, 137–148.
Suzuki, T., Hu, Z., Kumada, K., Phan, L.K., Zijlema, M., 2019. Non-hydrostatic modeling of drag, inertia and porous effects in wave propagation over dense vegetation fields. Coastal Eng. 149, 49–64.
Suzuki, T., Zijlema, M., Burger, B., Meijer, M.C., Narayan, S., 2012. Wave dissipation by vegetation with layer sschematisation in SWAN. Coastal Eng. 59 (1), 64–71.
Temmerman, S., Meire, P., Bouma, T.J., Herman, P.M., Ysebaert, T., De Vriend, H.J., 2013. Ecosystem-based coastal defence in the face of global change. Nature 504 (7478), 79–83.
Tinoco, R.O., Goldstein, E.B., Coco, G., 2015. A data-driven approach to develop physically sound predictors: application to depth-averaged velocities on flows through submerged arrays of rigid cylinders. Water Resour. Res. 51 (2), 1247–1263.
Tschirky, P., Hall, K., Turcke, D., 2001. Wave attenuation by emergent wetland vegetation. In: Coastal Eng., 2000, pp. 865–877.
Valentine, A., Kalnins, L., 2016. An introduction to learning algorithms and potential applications in geomorphometry and earth surface dynamics. Earth Surf. Dyn. 4 (2), 445–460.
van Rooijen, A.A., McCall, R.T., van Thiel de Vries, J.S.M., van Dongeren, A.R., Reniers, A.J.H.M., Roelvink, J.A., 2016. Modeling the effect of wave-vegetation interaction on wave setup. J. Geophys. Res. Oceans 121 (6), 4341–4359.
van Veelen, T.J., Fairchild, T.P., Reeve, D.E., Karunarathna, H., 2020. Experimental study on vegetation flexibility as control parameter for wave damping and velocity structure. Coastal Eng. 157, 103648.
van Veelen, T.J., Karunarathna, H., Reeve, D.E., 2021. Modelling wave attenuation by quasi-flexible coastal vegetation. Coastal Eng. 164, 103820.
van Wesenbeeck, B.K., Wolters, G., Antolínez, J.A., Kalloe, S.A., Hofland, B., de Boer, W. P., Bouma, T.J., 2022. Wave attenuation through forests under extreme conditions. Sci Rep 12 (1), 1884.
Vigen, T., 2015. Spurious Correlations. Hachette UK.
Wang, Y., Liu, Y., Yin, Z., Jiang, X., Yang, G., 2023. Numerical simulation of wave propagation through rigid vegetation and a predictive model of drag coefficient using an artificial neural network. Ocean Eng. 281, 114792.
Wang, Y., Yin, Z., Liu, Y., 2021. Predicting the bulk drag coefficient of flexible vegetation in wave flows based on a genetic programming algorithm. Ocean Eng. 223, 108694.
Williams, C.K., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning, 2. MIT Press, Cambridge, MA, p. 4.
Wu, W., Marsooli, R., 2012. A depth-averaged 2D shallow water model for breaking and non-breaking long waves affected by rigid vegetation. J. Hydraul. Res. 50 (6), 558–575.
Yang, S.L., Shi, B.W., Bouma, T.J., Ysebaert, T., Luo, X.X., 2012. Wave attenuation at a salt marsh margin: a case study of an exposed coast on the Yangtze Estuary. Estuaries Coasts 35, 169–182.
Ysebaert, T., Yang, S.L., Zhang, L., He, Q., Bouma, T.J., Herman, P.M., 2011. Wave attenuation by two contrasting ecosystem engineering salt marsh macrophytes in the intertidal pioneer zone. Wetlands 31, 1043–1054.
Zhang, K., Dresback, K.M., Irish, J.L., 2012. The reduction of storm surge by vegetation canopies. J. Geophys. Res. Oceans 117 (C3). https://doi.org/10.1029/2011JC0073804.
Zhang, X., Nepf, H., 2021. Wave damping by flexible marsh plants influenced by current. Phys. Rev. Fluids 6 (10), 100502.
Zhu, L., Chen, Q.J., Jafari, N., Rosati, J.D., Ding, Y., 2018. Modeling effects of vegetation on setup and runup of random waves. Coastal Eng. Proc. (36) 8-8.